BEST AVAILABLE IMAGES

Defective images within this document are accurate representation of The original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/02, 21/04, C07K 9/00
A61K 37/00, C12N 1/21, 5/10
C12N 15/70, C12P 21/00

A1

(11) International Publication Number: WO 94/10187

(43) International Publication Date: 11 May 1994 (11.05.94)

(21) International Application Number: PCT/US93/10307

(22) International Filing Date: 26 October 1993 (26.10.93)

(30) Priority data:
07/968,953 30 October 1992 (30.10.92) US
07/968,954 30 October 1992 (30.10.92) US

(60) Parent Applications or Grants
(63) Related by Continuation
US 07/968,953 (CIP)
Filed on 30 October 1992 (30.10.92)
US 07/968,954 (CIP)
Filed on 30 October 1992 (30.10.92)

(71) Applicants *(for all designated States except US)*: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP [CA/CA]; 88 Elm Street, Toronto, Ontario M5G 1X8 (CA). MEMORIAL SLOAN-KETTERING CANCER CENTER [US/US]; 1275 York Avenue, New York, NY 10021 (US). CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS [ES/ES]; Velazquez, 144, E-28006 Madrid (ES).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : LETARTE, Michelle [CA/CA]; 35 Howland Road, Toronto, Ontario M4K 2Z7 (CA). MASSAGUE, Joan [US/US]; 1233 York Avenue, New York, NY 10021 (US). BERNABEU, Carmelo [ES/ES]; Centro Investigaciones Biologicas, Velazquez, 144, E-28006 Madrid (ES). CHEIFETZ, Sela [CA/CA]; 88 New Seabury Drive, Concord, Ontario L4K 2R8 (CA).

(74) Agents: KONSKI, Antoinette, F. et al.; Campbell & Flores, 4370 La Jolla Village Drive, Suite 700, San Diego, CA 92122 (US).

(81) Designated States: AU, BB, BG, BR, CA, CZ, FI, HU, JP, KP, KR, LK, LV, MG, MN, MW, NO, NZ, PL, RO, RU, SD, SK, UA, US, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: COMPOSITIONS AND METHODS FOR MODIFYING THE REGULATORY ACTIVITY OF TGF-β

(57) Abstract

This invention provides a novel purified TGF-β binding glycoprotein, endoglin, and isolated nucleic acid molecules that encode amino acid sequences corresponding to the TGF-β-binding glycoprotein. Also provided is soluble endoglin-derived polypeptide, and fragments thereof. A pharmaceutical composition which comprises the purified endoglin-derived polypeptide or produced recombinantly methods and a pharmaceutically acceptable carrier is further provided as well as methods of treating patients by administering to the patient the pharmaceutical compositions of this invention.

COMPOSITIONS AND METHODS FOR MODIFYING THE REGULATORY ACTIVITY OF TGF-ß

FIELD OF THE INVENTION

The present invention relates to cell biology and to methods of modifying the biological activity of cell regulatory factors. More specifically, the present invention relates to a novel TGF-ß-binding glycoprotein.

Throughout this application various publications are referenced within parentheses. The disclosures of these publications in their entireties are hereby incorporated by reference to more fully describe the state of the art to which this invention pertains.

BACKGROUND OF THE INVENTION

Glycoproteins, in which one or more carbohydrate units have been attached covalently to the protein by posttranslational processing are widely distributed. Several secretory proteins, including the immunoglobulins, are glycoproteins, as are most components of plasma membranes such as cell membrane receptors, where the carbohydrates can be involved in cell-to-cell adhesion.

Transforming growth factor ß (TGF-ß) refers to a family of multi-functional cell regulatory factors produced in various forms by many cell types (for review see Sporn et al,. J. Cell Biol. 105:1039 (1987)). Five distinct isoforms of TGF-ß have been identified. TGF-ß1 and TGF-ß2 have been characterized in detail. TGF-ß is the subject of U.S. Patent Nos. 4,863,899; 4,816,561 and 4,742,003 which are incorporated herein by reference. TGF-ß binds to cell surface receptors present on various types of cells and is known to potentiate or inhibit the response of most cells to other growth factors, depending on the cell type. TGF-ß also regulates differentiation of some cell types, either promoting or inhibiting proliferation of the cell. Another

marked effect of TGF-ß is the promotion of cellular production of extracellular matrix proteins and their receptors (for a review see Keski-Oja et al., J. Cell Biochem. 33:95 (1987); Massague, Cell 49:437 (1987); Roberts and Sporn, "Peptides Growth Factors and Their Receptors", Springer-Verlag (1989)).

Notwithstanding the beneficial and essential cell regulatory functions served, TGF-ß regulatory activity can prove detrimental to its host organism. For example, whereas growth and proliferation of mesenchymal cells is stimulated by TGF-ß, some tumor cells may also be stimulated, using TGF-ß as an autocrine growth factor. In other cases the inhibition of cell proliferation by TGF-ß similarly proves detrimental to its host organism. An example would be the prevention of new cell growth to assist in repair of tissue damage. The stimulation of extracellular matrix production by TGF-ß is essential for wound healing. However, in some cases, the TGF-ß response is uncontrolled and an excessive accumulation of extracellular matrix results. An example of excessive accumulation of extracellular matrix is the "internal" scarring that occurs in the pathology glomerulonephritis and dermal scar tissue formation.

The transforming growth factor-β receptor system in most mesenchymal and epithelial cells consists of several components (Massague, J. Ann. Rev. Cell Biol. 6:597 (1990); Lin, H.Y. et al., Cell 68:775 (1992); Georgi, L.L. et al., Cell 61:635 (1990); Mathews, L.S. et al., Cell 65:973 (1991); Attisano, L. et al., Cell 68::97 (1992); Lopez-Casillas et al., Cell 67:785 (1991) and Wang et al., Cell 67:796 (1991)), one of which is betaglycan, a membrane-anchored proteoglycan. In addition to betaglycan, the TGF-β receptor system in most mesenchymal and epithelial cells consists of the type I receptor, a 53-kDa glycoprotein whose structure has not been determined yet,

and the type II receptor, which belongs to the protein serine/threonine kinase receptor family. Additional cell surface TGF-β-binding proteins, some of which have a more restricted distribution, have also been described.

Thus, a need exists to develop compounds that can modify the effects of cell regulatory factors such as TGF-ß. The present invention satisfies this need and provides related advantages as well.

SUMMARY OF THE INVENTION

The present invention provides a novel purified TGF-ß-binding glycoprotein. The protein, endoglin, is expressed at high levels on human vascular endothelial cells.

Further provided by the present invention are methods of treating pathologic conditions mediated by TGF-ß regulatory activity by contacting the TGF-ß with an effective amount of purified endoglin-derived polypeptide or any fragment thereof having the ability to bind TGF-ß. Thus, intact, native endoglin and soluble fragments thereof are useful in these methods. This invention provides a method of preparing and purifying full length and soluble endoglin-derived polypeptide. Isolated nucleic acids encoding the novel TGF-ß-binding glycoprotein and soluble endoglin-derived polypeptides are also provided, as well as vectors containing the nucleic acids and recombinant host cells transformed with such vectors.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows the domain structures of betaglycan and endoglin. Shown is a schematic representation highlighting regions of similarity between the linear sequences of betaglycan, an 853-amino acid

transmembrane proteoglycan, and endoglin, a disulfide-linked transmembrane protein composed of two identical subunits of 633 amino acids each. The transmembrane and short cytoplasmic regions (dark shaded box) of endoglin have a high level of sequence similarity to the corresponding regions of betaglycan. Two regions of weaker similarity are detected in the ectodomains of these proteins (light shaded boxes). Numbers represent the percent amino acid sequence similarity between the indicated domains of betaglycan and endoglin. Closed ovals represent positions of cysteine residues. Two putative sites for glycosaminoglycan chain attachment in betaglycan are indicated.

Figure 2 shows cell surface TGF-β1-binding proteins expressed by HUVEC. Near confluent cultures of HUVECs were affinity-labeled by incubation with 100 pM ^{125}I-TGF-β1 followed by chemical cross-linking with 0.16 mM disuccinimidyl suberate. A) Triton X-100 extracts of affinity-labeled HUVEC were resolved on SDS-PAGE gels under reducing (R) or nonreducing (NR) conditions. Lane C contains extract from cells affinity-labeled in the presence of excess unlabeled TGF-β1. The migration position of TGF-β receptors I (RI) and II (RII) are indicated. Arrow, the major affinity-labeled proteins of 180 kDa and higher molecular mass apparent on nonreducing gels. Arrowhead, the affinity-labeled proteins of 110-120 kDa seen on reducing gels. B) Detergent extracts of affinity-labeled HUVEC were resolved under nonreducing conditions on a first gel that was then resolved under reducing conditions in the second dimension as described in Cheifetz and Massague, J. Biol. Chem. 266:20767-20772 (1991), incorporated herein by reference. The 110-120-kDa labeled species migrating off-the-diagonal are indicated (arrowheads).

Figure 3 shows specific immunoprecipitation of TGF-β1-endoglin complexes. HUVECs were affinity-labeled with 100 pM ^{125}I-TGF-β1 as described in Figure 2. A) Detergent extracts of affinity-labeled cells were incubated with mAb 44G4 and immune complexes were collected on protein G-Sepharose. After washes, equal aliquots of the samples were analyzed under reducing (R) or non-reducing (NR) conditions by SDS-PAGE (5-8% polyacrylamide gradient gels). B) Affinity-labeled HUVEC lysates were maximally depleted of endoglin by two successive 45 min incubations at 4°C with 100 µl of 44G4-IgG-Sepharose. S) supernatant after second immunoprecipitation. I) the first 44G4 immunoprecipitation which contained 83% of the endoglin. T) corresponding amount of total extract used for the depletion experiment. All samples were analyzed under nonreducing conditions on SDS-PAGE with the exception of I_R, which was run under reducing conditions. The migration positions of TGF-β receptor II (RII), and endoglin monomer, dimer, and oligomer are indicated.

Figure 4 shows that endoglin transiently expressed in COS-M6 cells binds TGF-β1. COS-M6 cells were transfected with a cDNA encoding full-length L-endoglin (Endoglin) or control vector (C). Cells were affinity-labeled with 150 pM ^{125}I-TGF-β1 and the detergent extracts incubated with mAb 44G4 followed by protein G-Sepharose. Immunoprecipitated proteins were analyzed by SDS-PAGE under reducing (R) and non-reducing (NR) conditions and visualized by autoradiography.

Figure 5 shows the specificity of endoglin for TGF-β isoforms assessed in COS cell transfectants and in HUVEC. A) COS-M6 cells transfected with endoglin vector were affinity-labeled with 150 pM ^{125}I-TGF-β1 alone or in the presence of 1 or 10 nM unlabeled TGF-β1, -β2 or -β3. B) HUVEC were affinity-labeled with 100 pM ^{125}I-TGF-β1 alone or in the presence of 5 nM unlabeled TGF-β1 or TGF-β2.

Lysates from these cells were immunoprecipitated with MAb 44G4. Immunoprecipitates were fractionated under reducing conditions on SDS-PAGE gels. The region of the gels containing monomeric endoglin is shown along with the migration position of 100-kDa marker.

Figure 6 shows the restriction enzyme map of vector pcNeoSolEND with the 1.7 kb endoglin cDNA insert.

Figure 7 shows the partial nucleotide and predicted amino acid sequences of isolated S-endoglin cDNA. Nucleotides are numbered on the right. Amino acids are numbered on the left. The predicted signal sequence (nucleotides 283-357) and transmembrane region (nucleotides 2042-2116) are boldfaced. The 135 bp insert (underlined) contains splicing consensus sequences of donor/acceptor sites (GT, AG at positions 2134 and 2267) and branch point of lariat (CTGAC at position 2234). Nucleotides 372-2021 and 2322-3073 were found to be identical to the corresponding cDNA sequence of endothelial endoglin (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990) and Table 1).

Figure 8 shows the analysis of the cytoplasmic region which reveals the existence of two different forms of endoglin. A. Diagram of the cDNA coding for the cytoplasmic regions of the two alternative forms of endoglin. Only the region corresponding to the 3' end containing the cytoplasmic domain is depicted. The isolated endoglin cDNA (S-endoglin) contains a 135 bp insert not present in the previously described sequence (L-endoglin) (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990)). The sequence of the additional 135 bp insert is shown in Figure 7. The position of the stop codons and the corresponding translated protein sequences (thick bar), are indicated. B. Alignment of the cytoplasmic and transmembrane domains of S-endoglin with

the corresponding sequences of L-endoglin and human betaglycan, Morén, A. et al., Biochem. Biophys. Res. Commun. 189:356 (1992). Boxes contain identical sequences. Numbers indicate the position of the first amino acid in the whole sequence. Two major regions of identity were found. The first region (73% identity) involves residues 587-617 of S- and L-endoglin and residues 780-810 of betaglycan. The second region (74% identity), involves residues 634-660 of L-endoglin and residues 823-849 of betaglycan. The transmembrane region of S-endoglin is boldfaced. Dashes have been inserted for purposes of alignment. Asterisks indicate the last residue of the protein.

Figure 9 shows the expression of L-endoglin and S-endoglin in transfectant cells. Mouse fibroblasts were transfected with either L-endoglin or S-endoglin cDNA and the expression of the endoglin molecule analyzed. A. Analysis by cytofluorometry of the endoglin present at the cell surface. After trypsinization, cells were stained for indirect immunofluorescence with the monoclonal antibody 8E11 (anti-endoglin). A control staining of endoglin- mock transfectants is also shown. B. Immunoprecipitation analysis. Cells were metabolically labeled with [^{35}S]methionine, lysed and immunoprecipitated with 44G4 (anti-endoglin) or HC1/1 (anti CD11c) monoclonal antibody. Monoclonal antibody HC1/1 was included as a negative control. Samples were electrophoresed on a 6-12% acrylamide gradient gel under nonreducing conditions. C. Immunoblotting analysis. Mock (L cells), L-endoglin (L-Endo) and S-endoglin (S-Endo) cDNA transfected mouse fibroblasts and PMA-treated U937 cells were lysed in Triton X-100, and the insoluble material removed by centrifugation. Proteins contained in the supernatant were electrophoresed on a 6% acrylamide gel under non-reducing conditions and transferred to nitrocellulose membranes. Immunodetection of endoglin was carried out with 44G4

(anti-endoglin) monoclonal antibody using a chemiluminiscent assay. HC1/1 (anti-CD11c) and X63 monoclonal antibody were used as negative controls. D. Immunoprecipitation analysis of cell surface labeled endoglin. Mock, L-endoglin (L-Endo) and S-endoglin (S-Endo) transfected mouse fibroblasts were ^{125}I-labeled, lysed and immunoprecipitated with 44G4 monoclonal antibody (anti-endoglin). Samples were electrophoresed on a 10% acrylamide gel under either reducing (R) or nonreducing conditions (NR).

Figure 10 shows the detection of L-endoglin and S-endoglin transcripts by PCR amplification. Samples of total RNA from placenta, PMA treated HL-60 cells or PMA-treated U937 cells were incubated either in the presence (+RT) or in the absence (-RT) of reverse transcriptase. The generated cDNA samples, together with cDNA from S-endoglin (S-Endo) or L-endoglin (L-Endo) clones in pUC13 and cDNA from an endothelial cell library, were used for PCR amplification in the presence of oligonucleotides #14 and #15 specific for S-endoglin (panel A), or oligonucleotides #12 and #11 common to both L-endoglin and S-endoglin (panel B). No amplification was observed when the RT reaction was omitted, excluding the possibility of DNA contaminating the RNA samples. Additional bands below the specific amplified fragments of L and S endoglin probably represent primer-dimer artifacts.

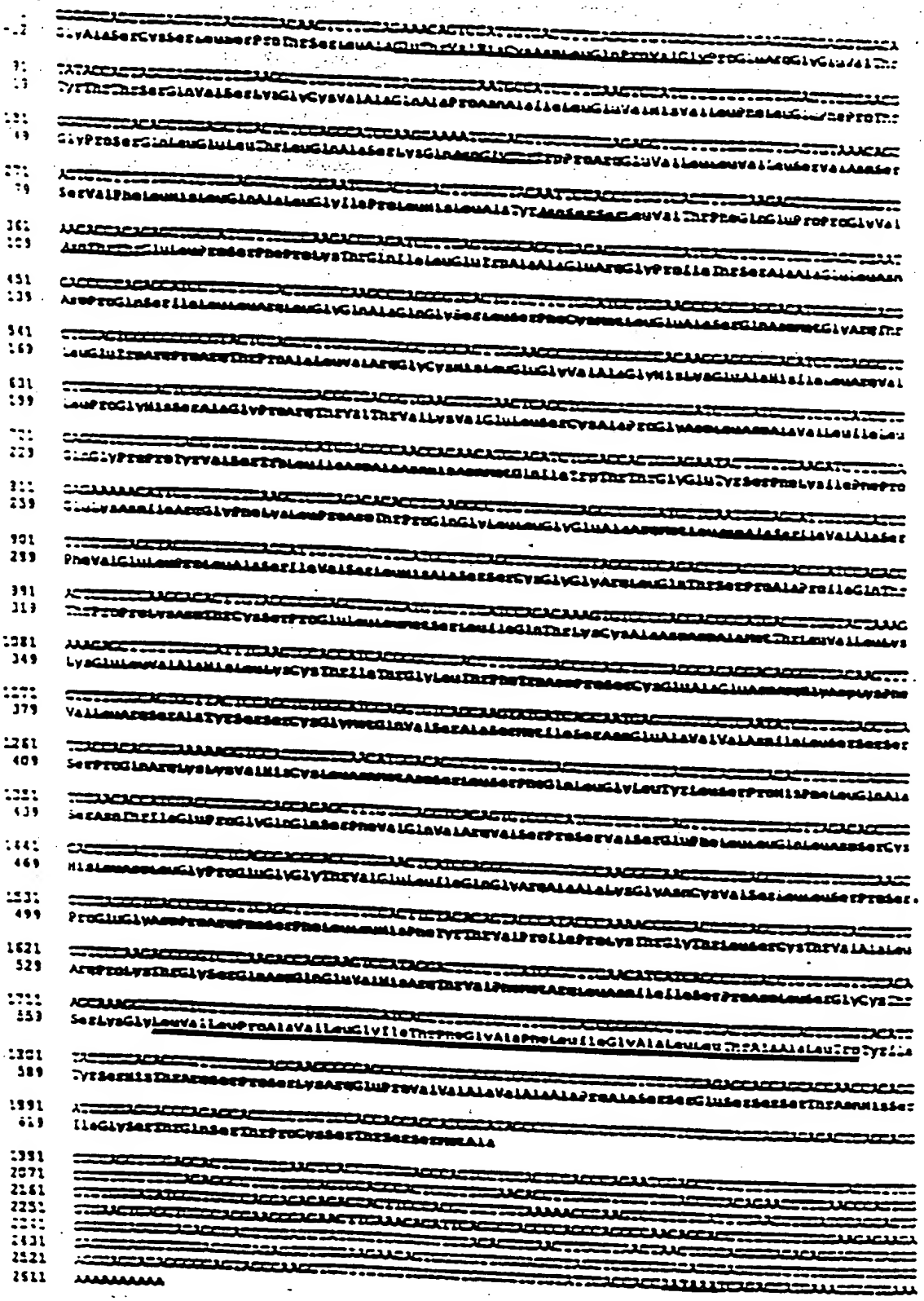
Figure 11 shows that both S-endoglin and L-endoglin bind TGF-ß1. Confluent cultures of S-endoglin (L+-S) and L-endoglin (L+-L) transfectants were affinity labeled by incubation with 100 pM ^{125}I-TGF-ß1 alone or in the presence of 4 nM unlabeled TGF-ß1, followed by chemical crosslinking with disuccinimidyl suberate. All samples bound specifically to TGF-ß1, as revealed by a fourfold ratio between the cpm bound in the absence versus presence of cold competing ligand; the parental L cells and mock

transfectants bound on average 60,000 specific cpm, while the endoglin transfectants bound on average 110,000 specific cpm. Immunoprecipitates of the L+-S transfectant with 44G4-IgG Sepharose contained 7700 cpm on average versus 700 cpm with control IgG-Sepharose; immunoprecipitates of the L+-L transfectant contained 4300 cpm versus 630 cpm for control. These immunoprecipitates were run on a 6-9% acrylamide gradient gel in reducing (R) or nonreducing (NR) conditions. The positions of the endoglin monomer, dimer and oligomer and of molecular weight markers are indicated.

Figure 12 shows the multiple cloning sites of pcDNAI/Neo.

Figure 13 shows the amino acid and nucleotide sequence of "L-Endoglin."

Table 1



DETAILED DESCRIPTION OF THE INVENTION

Endoglin is a homodimeric membrane glycoprotein composed of disulfide-linked subunits. Human-derived endoglin has been shown to exist in at least two isoforms expressed from a human cDNA library, an "L-isoform," of about 90 kDa or an "S-isoform" of about 85 kDa, reduced. Human endoglin, purified from tissue is shown to be composed of two disulfide-linked subunits each of about 95 kDa. It is expressed in human pre-erythroblasts, macrophages, leukemic cells of the lymphoid and myeloid lineages and at higher levels in vascular endothelial cells. It is also abundant on the syncytiotrophoblast, the multinucleated placental layer which constitutes the interface with maternal blood and plays an important role in providing nutrient exchange and immunological protection of the fetus (Gougos et al., Inter. Immunol. 4:83-92 (1992)).

Endoglin was first identified on a pre-B leukemic cell line, by its reactivity with mAb 44G4 (Quackenbush and Letarte, J. Immunol. 134:1276-1285 (1985)). It is present at low levels on cells derived from childhood acute leukemia cases (ALL) with pre-B lymphoid and myeloid phenotype; it is absent from T-ALL (Gougos and Letarte, J. Immunol. 141:1925-1933 (1988b); Kreindler et al., Leukemia and Lymphoma 3:7-18 (1990)). Interestingly, the human endoglin gene is localized to chromosome 9q34-qter, likely within the region translocated to chromosome 22 in Philadelphia chromosome-positive leukemia (Fernandez-Ruiz et al., Cytogen. Cell Gen. 64:204-207 (1993)). In normal adult bone marrow, only 3-5% of mononuclear cells express endoglin and they bear a pro-erythroblast phenotype; pre-B and myeloid precursors do not show detectable levels of endoglin (Buhring et al., Leukemia 5:841-847 (1991)). Normal B and T lymphocytes and unstimulated monocytes do not express endoglin but an up-regulation is observed on

activated macrophages (Lastres et al., Eur. J. Immunol. 22:393-397 (1992)).

Endoglin expression is considerably increased in the endothelium of various pathological skin lesions where endothelial cell proliferation is known to occur (Westphal et al., J. Invest. Dermatol. 100:27-34 (1993)). In tumors, capillary endothelial cells undergoing active angiogenesis also show higher levels of endoglin than resting endothelium in adjacent tissue.

Purified human endoglin exists in various isoforms and is composed of two disulfide linked subunits of Mr=95,000; 90,000; or 85,000, and bears N- and O-linked oligosaccharides. The primary sequence of human endoglin is composed of a 25 amino acid signal sequence, an extracellular domain of about 561 amino acids, a single transmembrane region of 25 amino acids and a cytoplasmic tail of 14 to 47 amino acid residues (Gougos and Letarte, J. Biol. Chem. 265:8361-8364 (1990)).

A relationship between human endoglin and the TGF-β receptor system was discovered with the molecular cloning of the rat TGF-β-binding proteoglycan, betaglycan (also known as the type III TGF-β receptor), which revealed that the transmembrane domain and the relatively short (43 amino acid) cytoplasmic tail of this protein were remarkably similar (71% amino acid sequence similarity and 63% amino acid identity) to the corresponding regions in endoglin (see Figure 1). The extracellular domains of these two proteins show limited homology in primary structure, and while endoglin is not a proteoglycan, it does contain N- and O-linked oligosaccharides.

Cloning of the TGF-ß receptor II (Lin et al., Cell 68:775-785 (1992)) revealed a functional transmembrane serine/threonine kinase which had previously been

identified as a polypeptide of 80 kd, when bound and chemically cross-linked to ^{125}I-TGF-ß (Cheifetz et al., J. Biol. Chem. 265:20533-20538 (1990); Cheifetz and Massagué, J. Biol. Chem. 266:20767-20772 (1991); Massagué, J., Cell 69:1067-1070 (1992)). This receptor associates with the type I receptor (a 53 kd protein) to transduce the TGF-ß signals; it is proposed that receptor I needs receptor II to bind TGF-ß and that receptor II needs receptor I to mediate a signal (Laiho et al., J. Biol. Chem. 266:9108-9112 (1991); Laiho et al., J. Biol. Chem. 265:18518-18524 (1990); Wrana et al., Cell 71:1003-1014 (1992)). Transfection experiments with a recently cloned receptor I, also a serine/threonine kinase, further supports the view that receptor I must associate with receptor II to bind TGF-ß (Ebner et al., Science 260:1344-1348 (1993)).

The present invention provides a soluble endoglin-derived human polypeptide that binds TGF-ß. The full-length soluble endoglin-derived polypeptide comprises a signal sequence that is cleaved during processing, the 561 amino acids of the extracellular domain of the mature endoglin polypeptide, an integral membrane protein, which consists of about 600 or 633 amino acids in total. Nucleic acid sequences encoding the human endoglin polypeptide are identified in Table 1 and Figure 13. The nucleic acid sequences encoding the soluble endoglin-derived polypeptides are included within the sequences set forth in Table 1 (from about amino acid number 1 to about amino acid number 561) and Figure 13. Also provided by this invention is a vector having inserted therein the genomic DNA molecule encoding endoglin. This vector was deposited under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure, with Coleccion Espanola de Cultivos Tipo (CECT), 46100 Burjasot (Valencia), Spain on October 21, 1993, under CECT 4475. Accordingly, isolated genomic DNA encoding endoglin is

within the scope of this invention.

This invention also provides purified human endoglin polypeptides encoding two isoforms differing from each other in the cytoplasmic region. The "S-endoglin" has 14 amino acid residues in the cytoplasmic region and the "L-endoglin" has 47 amino acid residues in the cytoplasmic region. The 586 amino acids spanning the extracellular and transmembrane regions in the mature endoglin are identical.

As used herein, the term "purified" means that the molecule or compound is substantially free of contaminants normally associated with a native or natural environment. For example, the mature human proteins can be obtained from a number of methods. The methods available for the purification of membrane proteins include precipitation, gel filtration, ion-exchange, reverse-phase, and affinity chromatography. Other well-known methods are described in Deutscher et al., Guide to Protein Purification: Methods in Enzymology Vol. 182, (Academic Press 1990), which is incorporated herein by reference. Alternatively, a purified polypeptide of the present invention can also be obtained by well-known recombinant methods as described, for example, in Sambrook et al., Molecular Cloning: A Laboratory Manual 2d ed. (Cold Spring Harbor Laboratory 1989), also incorporated herein by reference. An example of this means for preparing soluble endoglin-derived polypeptide is to express nucleic acid encoding the soluble endoglin in a suitable host cell, such as a bacterial, yeast or mammalian cell, using methods well known in the art, and recovering the expressed soluble protein, again using methods well known in the art.

For the purpose of illustration, expression of soluble endoglin was achieved by excising a full length 2.3 kb endoglin cDNA fragment from pcEXV-L ENDO, Hind III linkers were added and, after digestion with Hind III, it

was subcloned into the multiple cloning site of pBluescript vector (Stratagene), followed by digestion with Tth 111I. (This enzyme cuts the coding region approximately 80 bp upstream of the transmembrane region.)

Synthetic, complementary oligomers were engineered to contain an in-frame stop codon and a Bam HI overhang, and were ligated to the linearized plasmid. After digestion with Hind III, and Bam HI, and purification of the 1.7 kb fragment, it was ligated to pcDNAI/Neo (Invitrogen, San Diego, Figure 12). After cloning, the clones were screened by restriction enzyme analysis for the correct orientation and were found to be in the correct orientation. Verification of the cDNA insert was done using T7 and SP6 sequencing primers to confirm the presence of the start and stop codons at the 5′ and 3′ ends of the insert.

5′ End Primer

TAATACGACTCACTATAGGGAGACCCAAGCTTGGGGAATTCCGTGGACAGCATG

T7 PRIMER HINDIII ECORI Initiation

3′ End Primer

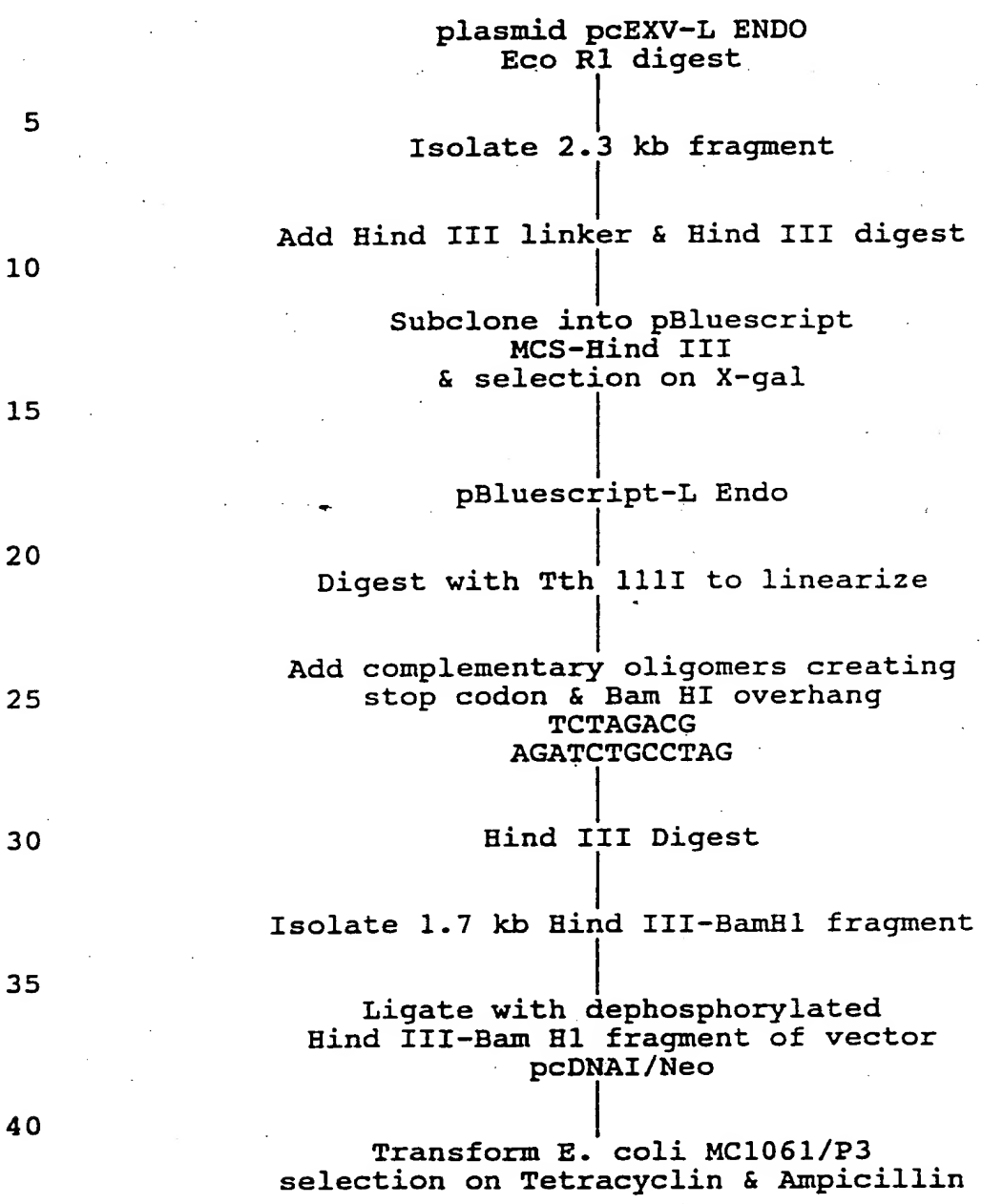
AAGACCGTCTAGACGGATCCACTAG.....CTATAGTGTCACCTAAATG

BAMHI SP6 PRIMER

To express the construct, CHO-k1 cells (ATCC) were transfected with pCDNA/Neo to determine optimal conditions required to generate stable transfectants resistant to G418. Five (5) x 10^6 cells electroplated with 5 μg of DNA at a voltage of 300 volts and a time constant of 17.9 msec., using the 960 uF capacitator (Biorad Gene Pulse) yielded the most number of stable transfectants with minimal cell death.

The following flow chart illustrates a means to express nucleic acid encoding soluble endoglin.

SOLUBLE ENDOGLIN



The soluble polypeptide and biologically active fragments thereof can also be produced by chemical synthesis. Synthetic polypeptides can be produced using Applied Biosystems, Inc. Model 430A or 431A automatic polypeptide synthesizer and chemistry provided by the manufacturer. The soluble polypeptide can also be isolated directly from cells which have been transformed with the expression vectors described below in more detail.

As used herein, endoglin-derived polypeptide means a human polypeptide having the amino acid sequence substantially the same as the 633 amino acid sequence shown in Table 1, or the 658 amino acid sequence shown in Figure 13, or an active fragment thereof. As used herein the term "soluble endoglin-derived polypeptide" refers to a soluble, biologically active fragment of the human endoglin polypeptide expressed by the extracellular domain of the nucleic acid. As used herein, an "active fragment" or "biologically-active fragment" refers to any portion of an endoglin polypeptide that binds to TGF-ß. Methods of determining whether a polypeptide can bind TGF-ß are well known to those of skill in the art, for example, as set forth herein.

The invention also encompasses nucleic acid molecules which differ from that of the nucleic acid molecule shown in Table 1, e.g., the sequence shown in Figure 13, but which produce the same phenotypic effect. These altered, but phenotypically equivalent nucleic acid molecules are referred to as "equivalent nucleic acids." This invention also encompasses nucleic acid molecules characterized by changes in non-coding regions that do not alter the phenotype of the polypeptide produced therefrom when compared to the nucleic acid molecule described above. This invention further encompasses nucleic acid molecules which hybridize to the nucleic acid molecule of the subject invention or its complement. As used herein, the term

"nucleic acid" encompasses mRNA and cRNA as well as single and double-stranded genomic DNA, DNA and cDNA. In addition, as used herein, the term "polypeptide" encompasses any naturally occurring allelic variant thereof, such as S-endoglin and L-endoglin, as well as man-made recombinant forms.

This invention provides an isolated nucleic acid molecule encoding a soluble endoglin-derived polypeptide. As used herein, the term "isolated nucleic acid molecule" means a nucleic acid molecule that is in a form that does not occur in nature. One means of isolating a human endoglin nucleic acid is to probe a human cDNA expression library with a natural or artificially designed antibody to endoglin, using methods well known in the art (see Gougos, A. et al., J. Biol Chem. 265:8361 (1990)) and the Examples set forth below. DNA and cDNA molecules which encode human endoglin polypeptides can be used to obtain complementary genomic DNA, cDNA or RNA from human, mammalian or other animal sources. Isolated genomic DNA also is encompassed by this invention as described above. It is isolated by using the nucleic acid sequences of this invention and methods well known to those of skill in the art as described in Sambrook et al., supra.

The invention further provides an isolated nucleic acid molecule operatively linked to a promoter of RNA transcription, as well as other regulatory sequences. As used herein, the term "operatively linked" means positioned in such a manner that the promoter will direct the transcription of RNA off of the nucleic acid molecule. Examples of such promoters are SP6, T4 and T7. Vectors which contain both a promoter and a cloning site into which an inserted piece of DNA is operatively linked to that promoter are well known in the art. Preferable, these vectors are capable of transcribing RNA in vitro or in vivo. Examples of such vectors are the pGEM series

(Promega Biotec, Madison, WI).

This invention provides an expression or replication vector comprising this isolated nucleic acid molecule such as DNA, cDNA or RNA encoding a soluble endoglin-derived polypeptide. Examples of vectors are viruses, such as bacteriophages, baculoviruses and retroviruses, cosmids, plasmids (such as pcEXV-2) and other recombination vectors. Nucleic acid molecules are inserted into vector genomes by methods well known in the art. For example, insert and vector DNA can both be exposed to a restriction enzyme to create complementary ends on both molecules that base pair with each other and which are then joined together with a ligase. Alternatively, synthetic nucleic acid linkers can be ligated to the insert DNA that correspond to a restriction site in the vector DNA, which is then digested with a restriction enzyme that recognizes a particular nucleotide sequence. Additionally, an oligonucleotide containing a termination codon and an appropriate restriction site can be ligated for insertion into a vector containing, for example, some or all of the following: a selectable marker gene, such as neomycin gene for selection of stable or transient transfectants in mammalian cells; enhancer/promoter sequences from the immediate early gene of human CMV for high levels of transcription; transcription termination and RNA processing signals from SV40 for mRNA stability; SV40 polyoma origins of replication and ColE1 for proper episomal replication; versatile multiple cloning sites; and T7 and SP6 RNA promoters for *in vitro* transcription of sense and anti-sense RNA. Other means are available.

Also provided are vectors comprising a DNA molecule encoding an endoglin-derived polypeptide, or soluble fragment thereof, adapted for expression in a bacterial cell, a yeast cell, a mammalian cell and other animal cells. The vectors additionally comprise the

regulatory elements necessary for expression of the DNA in the bacterial, yeast, mammalian or animal cells so located relative to the DNA encoding soluble endoglin polypeptide as to permit expression thereof. Regulatory elements required for expression include promoter sequences to bind RNA polymerase and transcription initiation sequences for ribosome binding. For example, a bacterial expression vector includes a promoter such as the lac promoter and for transcription initiation the Shine-Dalgarno sequence and the start codon AUG (Sambrook et al. supra. 1989). Similarly, a eucaryotic expression vector includes a heterologous or homologous promoter for RNA polymerase II, a downstream polyadenylation signal, the start codon AUG, and a termination codon for detachment of the ribosome. Such vectors can be obtained commercially or assembled by the sequences described in methods well known in the art, for example the methods described above for constructing vectors in general. Expression vectors are useful to produce cells that express the polypeptide.

This invention provides a mammalian cell containing a cDNA molecule encoding an endoglin-derived polypeptide or a soluble fragment thereof. An example is a mammalian cell comprising a plasmid adapted for expression in a mammalian cell. The plasmid has a cDNA molecule encoding an endoglin-derived polypeptide and the regulatory elements necessary for expression of the polypeptide. Various mammalian cells may be utilized as hosts, including, for example, mouse fibroblast cell NIH3T3, CHO cells, HeLa cells, Ltk- cells, etc. Expression plasmids such as those described supra can be used to transfect mammalian cells by methods well known in the art such as calcium phosphate precipitation, DEAE-dextran, electroporation or microinjection.

This invention provides a pharmaceutical composition containing a pharmaceutical carrier and any of

a purified polypeptide, a purified soluble polypeptide, an active fragment thereof, or a purified, mature protein and active fragments thereof, alone or in combination with each other. These polypeptides or proteins can be recombinantly derived, chemically synthesized or purified from native sources. As used herein, the term "pharmaceutically acceptable carrier" encompasses any of the standard pharmaceutical carriers, such as a phosphate buffered saline solution, water and emulsions such as an oil/water or water/oil emulsion, and various types of wetting agents. Any of these pharmaceutical compositions are useful in the methods described below or for the preparation of a medicament for treating the conditions described below.

Also provided are antibodies having specific reactivity with the endoglin-derived TGF-ß-binding polypeptides of the subject invention, such as anti-endoglin antibody 44G4 (Quackenbush, E.J., and Letarte, M.J., J. Immunol. 134:1276-1285 (1985)) or any antibody having specific reactivity to a TGF-β-binding endoglin polypeptide. Active fragments of antibodies are encompassed within the definition of "antibody." The antibodies and fragments of the invention can be produced by any method known in the art. For example, polyclonal and monoclonal antibodies can be produced by methods well known in the art, as described, for example, in Harlow and Lane, Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory 1988), which is incorporated herein by reference. The polypeptide can be used as the immunogen in generating such antibodies. Altered antibodies, such as chimeric, humanized, CDR-grafted or bifunctional antibodies can also be produced by methods well known to those skilled in the art. Such antibodies can also be produced by hybridoma, chemical synthesis or recombinant methods described, for example, in Sambrook et al., supra. The antibodies can be used for determining the presence or purification of the endoglin-derived polypeptide or soluble

fragment thereof, of the present invention. With respect to the detecting of such polypeptides, the antibodies can be used for in vitro diagnostic methods to determine the presence of endoglin or in vivo imaging methods.

Immunological procedures useful for in vitro detection of the target soluble endoglin-derived polypeptide in a sample include immunoassays that employ a detectable antibody. Such immunoassays include, for example, ELISA, Pandex microfluorimetric assay, agglutination assays, radioimmunoassays, flow cytometry, serum diagnostic assays and immunohistochemical staining procedures which are well known in the art. An antibody can be made detectable by various means well known in the art. For example, a detectable marker can be directly or indirectly attached to the antibody. Useful markers include, for example, radionuclides, enzymes, fluorogens, chromogens and chemiluminescent labels.

This invention provides a method of modifying a biological function mediated by the regulatory activity of TGF-ß which comprises contacting a suitable sample containing TGF-ß with an effective amount of a biologically active endoglin-derived polypeptide, for example soluble endoglin, or a pharmaceutical composition described above.

As used herein, "an effective amount" refers to an amount of the polypeptide sufficient to bind to TGF-ß and thereby prevent or inhibit its regulatory activity. This method is especially useful for modifying the regulatory activity of TGF-ß1 or TGF-ß3. Examples of regulatory activities include, but are not limited to stimulation of cell proliferation, cell growth inhibition, promotion of extracellular matrix proteins, and regulation of immune functions. TGF-ß is known to be a potent chemoattractant for monocytes and can induce IL-1, TNF-α, TGF-ß and surface FcγRIII, all of which are involved in the

inflammatory response. Conversely, TGF-ß can deactivate macrophages by inhibiting the antimicrobial activity and the superoxide anion generation, and induce suppression of class II-restricted Ag presentation by macrophages.

The method can be practiced in vitro or in vivo. If the method is practiced in vitro, contacting is effected by incubating the sample with a polypeptide, a protein or a pharmaceutical composition as described above.

In vitro the novel nucleic acid molecules and antibodies of this invention are useful to detect and quantify the amount of TGF-ß in a sample isolated from a subject, such as a human patient. The detection of TGF-ß is useful to monitor the progression of a disease related to overexpression of TGF-ß, e.g., glomerulonephritis.

However, in a preferred embodiment the contacting is effected in vivo by administering a polypeptide, a protein or a pharmaceutical composition, as described above, to a subject, e.g., a human patient.

Methods of administration are well known to those of skill in the art and include, but are not limited to administration orally, intravenously or parenterally. Administration will be in such a dosage such that the regulatory activity is effectively modified. Administration can be effected continuously or intermittently such that this amount is effective for its intended purpose.

This invention also provides a method of treating a pathologic condition caused by a TGF-ß-regulated activity comprising contacting the TGF-ß with any of a purified soluble endoglin-derived polypeptide, an active fragment thereof, an endoglin-derived polypeptide or an active fragment thereof. The TGF-ß is bound with said polypeptide

to thereby treat the pathologic condition mediated by TGF-ß regulatory activity. As used herein, "pathologic conditions" refers to any pathology arising from TGF-ß-induced regulatory activity, for example, inflammation, rheumatoid arthritis, inflamed skin lesions, scar tissue formation, lung fibrosis, liver fibrosis, atherosclerosis, and glomerulonephritis. Growth and proliferation of mesenchymal cells is stimulated by TGF-ß, however some tumor cells may also be stimulated thus using TGF-ß as an autocrine growth factor. An example of inhibitory conditions are the prevention of new cell growth to assist in repair of tissue damage, for ulceration and immunosuppression. The stimulation of extracellular matrix production by TGF-ß is essential for wound healing. However, in some cases, the TGF-ß response is uncontrolled and an excessive accumulation of extracellular matrix results. An example of excessive accumulation of extracellular matrix is glomerulonephritis. An additional example of a pathology is cancer.

In one embodiment, the method is practiced by administering to a subject, e.g., a human patient or a mammal, an effective amount of a purified endoglin protein or an endoglin-derived soluble polypeptide or a biologically active fragment thereof, or the pharmaceutical composition described above. Methods of administration are outlined supra.

Also provided by this invention is a method of inhibiting the activity of endoglin by contacting endoglin with an effective amount of a polypeptide capable of binding to endoglin to bind endoglin, thereby inhibiting the activity of endoglin. As used herein, the term "a polypeptide capable of binding to endoglin" means any substance capable of forming a complex with endoglin, for example, TGF-ß1 or TGF-ß3, or an active fragment thereof. An active fragment is an amino acid sequence corresponding

to a fragment of TGF-ß1 or TGF-ß3 that retains the ability to bind endoglin. Methods of making such fragments are well known to those of skill in the art as are methods of determining the binding activity of the fragments. Also encompassed by this invention are polypeptides that retain their activity to bind to endoglin, but no longer mediate the biological response corresponding to the binding of a functional ligand to the receptor is destroyed. Thus, these "mutated" polypeptides can act as antagonists to the biological function mediated by the ligand to endoglin by blocking the binding of normal, functioning ligands to endoglin on the cell.

This invention also encompasses the use of the compositions defined above for the preparation of medicaments to modify a biological function regulated by TGF-ß. These biological functions are described above in detail.

It is understood that modifications which do not substantially affect the activity of the various molecules of this invention are also included within the definition of said molecules.

The following examples are intended to illustrate but not limit the present invention.

A. ISOLATION OF HUMAN ENDOGLIN PROTEIN AND NUCLEIC ACID ENCODING SAME

EXAMPLE I

Cell Culture and Transfections

Human umbilical vein endothelial cells (HUVEC, CRL 1730, ATCC) were maintained in α-minimal essential media supplemented according to supplier's instructions or prepared from umbilical veins as previously described

(Gougos, A. et al., J. Immunol. 141:1925 (1988)). Similar results were obtained using cells from either source. COS-M6 cells, maintained in Dulbecco's modified Eagle's medium supplemented with 10% bovine serum, were transfected with a cDNA encoding full-length endoglin ligated into the EcoRI site of the mammalian expression vector pcEXV (Miller, J. et al., J. Exp. Med. 164:1478 (1986)) or with a control vector without cDNA insert (pcMV5; Lopez-Casillas, F. et al., Cell 67:785 (1991)) by the DEAE-dextran-chloroquine procedure (Seed, B., et al., P.N.A.S. USA 84:3365 (1987)). Twenty-four (24) hours post-transfection, cells were trypsinized and reseeded into multicluster dishes and allowed to grow an additional 48 hours before being affinity-labeled with ^{125}I-TGF-β1 as described below.

EXAMPLE II

Receptor Affinity Labeling and Immunoprecipitation

TGF-β1 and TGF-β2 were purchased from R & D Systems (Minneapolis, MN) and TGF-β3 was obtained from Oncogene Science (Manhassett, NY). ^{125}I-TGF-β1 used in these studies was prepared by the chloramine-T method as previously described (Cheifetz, S. et al., J. Biol. Chem. 265:20533 (1990)) or purchased from Amersham Corp.; both preparations gave identical results. The conditions for affinity labeling cell monolayers with ^{125}I-TGF-β1 and disuccinimidyl suberate (Pierce Chemical Co.) have been described previously (Massague, J., Methods Enzymol. 146:174 (1987)). The concentrations of ^{125}I-TGF-β1 and competing unlabeled ligands used for each experiment are indicted in the figure legends. Triton X-100 extracts of the affinity-labeled cells were either analyzed directly on sodium dodecyl sulfate-polyacrylamide gels (SDS-PAGE) or first incubated with monoclonal antibody (mAb) 44G4 directed against human endoglin (Quackenbush, E.J. et al., J. Immunol. 134:1276 (1985)) or with control antibody (see below). For immunoprecipitations, detergent extracts were

diluted with an equal volume of phosphate-buffered saline containing 1% Triton X-100 and precleared by incubation for 20 min at 4°C with protein G-Sepharose (Pharmacia LKB Biotechnology Inc.) prior to overnight incubation at 4° C with mAb 44G4. Immune complexes were collected by incubation with protein G-Sepharose for 1 hour at 4° C. For some experiments, mAb 44G4 was used coupled to Sepharose. The immunoprecipitates were washed three times (saline with 1% Triton X-100) and then resolved by SDS-PAGE in the presence or absence of dithiothreitol (DTT) and visualized by autoradiography. Irrelevant mAb (44D7) used in control experiments to monitor specificity of the immunoprecipitations did not immunoprecipitate any affinity-labeled bands.

EXAMPLE III

SDS-PAGE and 2D-Gel Analysis

Analysis of the affinity-labeled profile of HUVEC revealed that, like vascular endothelial cells from other sources, these cells have little or no betaglycan, which characteristically migrates as a diffuse band between 200 and 400 KDa on reducing SDS-PAGE (Figure 2A). Instead, HUVEC expressed a disulfide-linked cell surface protein that, together with TGF-β receptors I and II, was affinity-labeled by crosslinking with ^{125}I-TGF-β1. Receptors I and II were detected in HUVEC as labeled complexes of approximately 65 KDa and 100 KDa, respectively, which is similar to the size of these labeled receptors reported for other human cell lines. Comparison of the relative migration of the affinity-labeled proteins fractioned on SDS-PAGE revealed that the major affinity-labeled proteins of HUVEC migrated between 95-120 KDa on reducing gels whereas on non-reducing gels the major affinity-labeled proteins migrated between 100-110 KDa (presumed to be receptor II) and at 180 KDa and above (endoglin) (Figure 2). This pattern indicated the presence of disulfide-

linked TGF-β-binding proteins.

Resolution of these disulfide-linked TGF-β1 binding proteins on two-dimensional gels (Figure 2B) confirmed that the disulfide-linked complexes (probably dimers and higher order oligomers) contained subunits of approximately 95 KDa (value estimated by subtracting the cross-linked TGF-β1 monomer mass 12.5 KDa from the reduced 110 KDa affinity-labeled complex). Together with the type II receptor, the disulfide-linked TGF-ß1-binding proteins are the major affinity-labeled species expressed by HUVEC.

EXAMPLE IV

Immunoprecipitation with anti-endoglin mAb

To determine whether the disulfide-linked TGF-β-binding protein on endothelial cells was endoglin, affinity-labeled HUVEC extracts were immunoprecipitated with monoclonal antibody (mAb) 44G4 which is specific for human endoglin (Georgi, L.L. et al., Cell 61:635 (1990); MacKay, K. et al., J. Biol. Chem. 266:9907 (1992); Merwin, J.R. et al., Am. J. Pathol. 138:37 (1991)). Electrophoretic analysis of these immunoprecipitates revealed a labeled protein complex whose subunit structure was similar to that of endoglin (Figure 3A). Thus, under reducing conditions, a major affinity-labeled band of approximately 110 KDa was seen which migrated as complexes of 180 KDa and greater than 200 KDa when analyzed under non-reducing conditions. The higher order oligomers might contain multiple endoglin molecules crosslinked by TGF-β1, itself a disulfide-linked dimer. Repeated immunoprecipitation with 44G4-IgG-Sepharose completely depleted these labeled species from cell extracts (Figure 3B). No affinity-labeled bands were immunoprecipitated from three other human cell lines (A549, Hep G2, MCF-7), which lack endoglin and were used as negative-controls for these experiments. Monoclonal antibodies specific to human

endoglin thus demonstrate that endoglin is a major TGF-β-binding protein in human vascular endothelial cells.

EXAMPLE V

Ectopic Expression of Endoglin in Cells

The identity of this dimeric TGF-β-binding protein of HUVEC with endoglin was confirmed by ectopically expressing the full-length endoglin cDNA in COS monkey kidney cells. After affinity-labeling with ^{125}I-TGF-β1, a labeled species with the characteristics of endoglin could be specifically precipitated by mAb 44G4 only from the detergent extracts of endoglin transfectants (Figure 4). Differences in glycosylation likely account for the smaller size of endoglin expressed in COS cells relative to endogenous endoglin of HUVEC.

B. ISOLATION OF S- AND L-ENDOGLIN

EXAMPLE VI

cDNA Cloning and Sequencing of Endoglin

Approximately 2.5 X 10^5 clones from a λgt10 cDNA library prepared from PMA-treated myelomonocytic human cell line HL60 (Corbí, A.L. et al., EMBO J. 6:4023 (1987)). (50,000 pfu/150 mm dish) were screened with 700 bp PstI fragment from endoglin cDNA (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990)). Thirteen hybridizing clones were isolated after three rounds of plaque purification with sizes ranging from 2.4-3.1 kb. The longest clone (clone 3.3; 3073 bp) was subcloned into pUC13 and sequenced using the dideoxy chain termination method (Sanger, F. et al., PNAS USA 74:5463 (1977)).

Clone 3.3 in pUC13 was digested with BbrPI and BamHI. Endoglin fragment was made blunt and inserted into the mammalian expression vector pcEXV (Miller, J. and

Germain, R., J. Exp. Med. 164:1478 (1986)), yielding pcEXV-EndoS. The lack of leader sequence in the cDNA (Table I) (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990) was overcome by the construction of pcEXV-EndoL. pcEXV-EndoS was digested with MluI/BamHI and ligated to the 563 bp MluI/BamHI fragment specific to endoglin cDNA (Table I), resulting in pcEXV-EndoL. Transfectants were generated by cotransfection of either pcEXV-EndoS or pcEXV-EndoL with psV2neo into mouse L cells. After G418 selection, endoglin-positive clones were isolated. In the same experiment, endoglin-negative clones were selected as mock transfectants.

EXAMPLE VII

Antibodies, Immunofluorescence, Immunoprecipitations and Immunoblotting

Endoglin-specific monoclonal antibodies used were 8E11 (Lastres, P. et al., Eur. J. Immunol. 22:393 (1992)) of the IgM class and 44G4 (Gougos, A. and Letarte, M., J. Immunol. 141:1925 (1988)) of the IgG1 subclass. Control monoclonal antibodies were HC1/1 (anti-CD11c) of the IgG1 subclass and X63. FCM and immunoprecipitation analyses were performed as described in Lastres, P. et al., Eur. J. Immunol. 22:393 (1992), incorporated herein by reference. For immunoblotting studies, 5 X 10^6 cells were lysed in 250 μl of 1% Triton X-100, 1 mM PMSF in PBS for 30 minutes at 4°C. Insoluble material was removed by centrifugation at 100,000 x g for 1 hour in a Beckman TL100 centrifuge. Proteins contained in the supernatant were separated by SDS-PAGE under nonreducing conditions using a minigel system and transferred to nitrocellulose membranes (Hybond, Amersham). Membranes were incubated first with a blocking solution (10% fetal calf serum, 0.5% Tween-20, 1 M glucose and 10% glycerol in PBS), followed by incubation with the monoclonal antibody 44G4. The presence of endoglin was revealed using a chemiluminescence assay (ECL detection

kit, Amersham).

EXAMPLE VIII

Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

Total cellular RNA was isolated using guanidinium isothiocyanate and a cesium chloride ultracentrifugation step (Chirgwin, J.M. et al., Biochemistry 18:5294 (1979)). Poly (A)+ RNA was purified by oligo (dT) affinity chromatography (Maniatis, T. et al., "Molecular Cloning: A Laboratory Manual" Cold Spring Harbor Laboratory, Cold Spring Harbor (1982). Single-stranded cDNA was synthesized from poly (A+) RNA (0.5 μg/40 μl reaction), using AMV reverse transcriptase. Five microliters of cDNA was used for a 50 μl PCR reaction (Gilliland, G. et al., In: [Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (Eds.)) "PCR Protocols" Academic Press, San Diego 1990, p. 60.). Oligonucleotide primers used for amplification were E#12 (nucleotides 1945-1964), E#11R (reverse complement of nucleotides 2475-2494), E#14 (nucleotides 2136-2158), and E#15 (reverse complement of nucleotides 2247-2273). Amplifications were performed in 1 X Taq buffer (Promega) with 0.2 mM each dATP, dCTP, dGTP, dTTP at 0.25 μM concentration of each primer and 0.25 U/50 μl of Taq DNA polymerase (Promega). Amplification was carried out in a thermal cycler as follows: 5 minutes at 95°C; 35 cycles of 45 seconds at 94°C, 45 seconds at 54°C, and 1 minute at 72°C for oligonucleotide pair E#12-E#11R or 35 cycles of 45 seconds at 94°C, 45 seconds at 68°C and 1 minute at 72°C for pair E#14-E#15 and then 10 minutes at 72°C. Control reactions, identical except for the omission of reverse transcriptase, were performed simultaneously. Amplification products were analyzed by agarose gel electrophoresis and ethidium bromide staining.

EXAMPLE IX

Receptor Affinity Labeling

TGF-ß1 was purchased from R&D Systems (Minneapolis, MN) and ^{125}I-TGF-ß1 was obtained from Amersham (Oakville, Canada) at a specific activity of 2000 Ci/mmol. Parental mouse L cells, transfectant mouse L cells expressing S-endoglin or L-endoglin, and mock transfectant L cells were grown at confluence (5 X 10^6 cells per plate) and incubated with 100 pM of ^{125}I-TGF-ß1 for 4 hours with and without 4 nM of competing unlabeled TGF-ß1. Cells were washed and cross-linked at a final concentration of 0.16 mM disuccinimidyl suberate (DSS) (Pierce Chemical Co.) in a buffer containing 128 mM NaCl, 5 mM KCl, 5 mM $MgSO_4$, 1.2 mM $CaCl_2$, 50 mM Hepes, pH 7.5 for 15 minutes at 4°C. Cells were washed four times and solubilized directly on the petri dishes with a minimum of solubilization buffer containing 1% Triton X-100 and proteolytic inhibitors as described previously (Massagué, J., Methods Enzymol. 146:174 (1987)). The extracts were subjected to immunoprecipitation followed by SDS-PAGE analysis.

RESULTS

Isolation and Characterization of Endoglin cDNA Clones: Identification of a Cytoplasmic Variant

Full-length cDNA clones were derived from a λgt10 library prepared from PMA-treated HL60 cells. Clone number 3.3 was selected for further characterization due to the large size of the insert. The partial sequence of the 3073 bp cDNA insert from clone 3.3 is depicted in Figure 7. This clone contains 281 bp of 5′ untranslated region followed by an open reading frame of 1875 bp. The predicted protein sequence shows that the endoglin leader peptide contains 25 amino acids (aa), followed by 561 residues at the extracellular portion and a transmembrane

region spanning 25 amino acids as expected (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990)). However, clone 3.3 contains a 135 bp segment inserted within the previously known cDNA sequence, starting at nucleotide (nt) 2134 (Figure 8). The first 21 nucleotides of the insert are in frame with the preceding sequence and code for a new sequence of 7 amino acids. This sequence replaces the C-terminal 40 amino acids of endoglin, thus leading to a new cytoplasmic tail of 14 amino acids different from the 47 residues previously reported (Gougos, A. and Letarte, M.J., J. Biol. Chem. 265:8361 (1990)) (Figure 8), suggesting the existence of two alternative forms of endoglin. The predominant form of endoglin in the myelomonocytic cell line HL60 seems to be the one containing 47 residues in the cytoplasmic domain since this sequence was present in 12 out of 13 clones analyzed. The isoforms with the 14 and 47 amino acid cytoplasmic tails, and the corresponding cDNAs, are referred to as S-endoglin and L-endoglin, respectively.

Expression of two different forms of endoglin

To characterize further the two alternative forms, independent constructs corresponding to the short and long forms of endoglin were inserted into the expression vector pcEXV and transfected into mouse fibroblasts (Figure 9). Both, S-endoglin and L-endoglin were highly expressed on the cell surface as determined by FCM (Figure 9). In addition, metabolic labeling of the transfectants followed by immunoprecipitation, revealed a 170-kDa (L-endoglin) or a 160-kDa (S-endoglin) protein specifically recognized by anti-endoglin monoclonal antibody (Figure 9). The distinct size of S- and L-endoglin in the transfectants was also detected by immunoblotting analysis under non-reducing conditions (Figure 9). In these experiments, L-endoglin showed the same M_r as the endoglin detected on U937 cells, suggesting that this is the predominant form in the promonocytic cell

line.

Cell surface radiolabeling of the transfectants followed by immunoprecipitation demonstrated that both isoforms are expressed as disulfide linked homodimers (Figure 9), indicating that cysteine residue present in the extracellular portion are mediating the interchain disulfide bond.

Differential expression of S-endoglin and L-endoglin mRNA

The individual expression of S- and L-endoglin mRNA was analyzed by RT-PCR. Amplification with two primers derived from the unique 3' region of S-endoglin cDNA generated the expected 137 bp product on S-endoglin transfectants and on PMA-treated monocytic lines, endothelial cells and placenta (Figure 10), demonstrating the existence of this isoform in several cell types. When primers common to L-endoglin and S-endoglin were used, the L-endoglin specific fragment of 411 bp could be amplified on PMA-treated monocytic cell lines, placenta and endothelial cells, whereas the S-endoglin fragment of 546 bp was only amplified on control S-endoglin transfectants (Figure 10). When competitive templates are amplified by PCR, an abundant template can suppress the amplification of a less abundant one (Gilliland, G. et al., In: [Innis, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (Eds.)) "PCR Protocols", Academic Press, San Diego 1990, p. 60). Therefore, these results indicate that L-endoglin is the predominant form and S-endoglin is expressed at lower levels in these cell types. This is in agreement with the presence in PMA-treated U937 cells of an endoglin molecule of M_r similar to that observed on the transfectant L-endoglin (Figure 9), assuming similar glycosylation levels on both cell types.

L-endoglin and S-endoglin bind to TGF-ß1

Since endoglin has been found to be a component of the receptor system for TGF-ß, Cheifetz, S. et al., J. Biol. Chem. 267:19027 (1992), it was of interest to analyze the ability of each isoform to bind this ligand. Transfectants expressing short and long forms of endoglin were compared to the parental L cells and to the mock transfectants for their ability to bind TGF-ß1. Immunoprecipitation analysis allowed the identification of the TGF-ß1-endoglin complex (Figure 11). Under non-reducing conditions, the dimers of endoglin were seen as radiolabelled bands of M_r 170 kDa (short form) and 175 kDa (long form). An oligomer of endoglin is seen in both cases migrating with a M_r > 270 kDa. A similar complex was also observed previously in cross-linking experiments with human endothelial cells and might represent endoglin molecules cross-linked by TGF-ß1, itself a dimer (Roberts, A.B. and Sporns, M.B. In: Sporn, M.B. and Roberts, A.B. (Eds.), "Peptide growth factors and their receptors" Springer-Verlag, Heidelberg 1990, p. 419). Under reducing conditions, major bands of 97 kDa and 107 kDa were seen with the short form of endoglin and bands of 102 kDa and 112 kDa were observed with the long form. These doublets have been observed previously for endothelial cells and might represent endoglin bound to the monomer (12.5 kDa) or dimer (25 kDa) of TGF-ß1 (Roberts, A.B. and Sporns, M.B. In: Sporn, M.B. and Roberts, A.B. (Eds.), "Peptide growth factors and their receptors" Springer-Verlag, Heidelberg 1990, p. 419.; Cheifetz, S. and Massagué, J., J. Biol. Chem. 266:20767 (1991)). By subtracting the contribution of TGF-ß1, one can estimate a molecular weight of 85 kDa for S-endoglin and 90 kDa for L-endoglin.

DISCUSSION

Sequence analysis demonstrated the existence of two different cDNA variants named L-endoglin and S-endoglin. These two isoforms are coexpressed by myeloid cells, endothelium and placenta, although the majority of the transcripts synthesized apparently correspond to the L-endoglin isoform. The mechanism by which the two isoforms are generated remains to be determined. Most likely, both isoforms are generated by alternative splicing. In fact, consensus sequences of donor/acceptor sites (GT, AG) at the 5′ and 3′ ends and branch point of the lariat (CTGAC), have been found on the novel 135 bp insert of S-endoglin cDNA (Figure 7). A similar example of cytoplasmic variants generated by a "retained intron" splicing mechanism, have been recently reported for the activin receptor of the TGF-ß receptor family (Attisano, L. et al., Cell 68:97 (1992)).

The behavior of the two endoglin isoforms was analyzed by transfection studies. Both forms are expressed on the cell surface as disulfide linked homodimers, indicating that the cysteine residues present in the extracellular region are responsible for the dimerization. In spite of the different cytoplasmic domains, both forms behave as TGF-ß1 binding proteins.

S-endoglin, L-endoglin and betaglycan contain 22 identical residues in the transmembrane domain and the adjacent cytoplasmic region (Figure 9). This delineates a first conserved motif between betaglycan and the two forms of endoglin, which contains two tyrosine residues and two Ser/Thr residues whose phosphorylation status remain to be analyzed. A second region of high identity in the cytoplasmic tail is shared only by the long form of endoglin and betaglycan. The high content (40%) of Ser/Thr residues in this motif, and the absence of Tyr suggests that this region might undergo phosphorylation by a Ser/Thr

kinase. Interestingly, the cytoplasmic domain of the TGF-β receptor II, displays Ser/Thr kinase activity (Lin, H.Y. et al., Cell 68:775 (1992)).

Although the invention has been described with reference to the disclosed embodiments, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.

What is claimed is:

1. An isolated nucleic acid molecule encoding an endoglin polypeptide or a biologically active fragment thereof.

2. A nucleic acid molecule encoding a soluble endoglin-derived polypeptide or a biologically active fragment thereof.

3. The nucleic acid of claim 1 or 2, wherein the nucleic acid encodes a human endoglin polypeptide.

4. The nucleic acid molecule of claim 1 or 2, wherein the nucleic acid is genomic DNA, cDNA, mRNA or cRNA.

5. The nucleic acid molecule of claim 4, wherein the genomic DNA is designated CECT 4475.

6. The nucleic acid molecule of claim 1 or 2 operatively linked to a promoter of RNA transcription.

7. A vector containing the nucleic acid molecule of claim 5.

8. A host cell containing the vector of claim 7.

9. The host cell of claim 8, wherein the cell is a procaryotic cell or a eucaryotic cell.

10. A method of preparing an endoglin-derived polypeptide or active fragment thereof, comprising:

a. inserting a nucleic acid molecule encoding an endoglin-derived polypeptide or active fragment thereof into a suitable expression vector;

b. inserting the resulting vector into a suitable host cell;

c. inducing the resulting host cell to express the endoglin-derived polypeptide or active fragment thereof; and

d. purifying the resulting endoglin-derived polypeptide so produced.

11. Soluble endoglin-derived polypeptide.

12. Endoglin-derived polypeptide or a biologically active fragment thereof produced by the method of claim 10.

13. A pharmaceutical composition comprising the polypeptide of claim 11 or 12 and a pharmaceutically acceptable carrier.

14. A pharmaceutical composition comprising a pharmaceutically acceptable carrier and a purified and isolated human endoglin-derived polypeptide or a biologically active fragment thereof substantially free of other host cell proteins.

15. A method of modifying a biological function mediated by the regulatory activity of TGF-ß which comprises contacting TGF-ß with an effective amount of an endoglin-derived polypeptide or a biologically active fragment thereof, thereby modifying the biological function.

16. The method of claim 15, wherein TGF-ß is TGF-ß1.

17. The method of claim 15, wherein TGF-ß is TGF-ß3.

18. The method of claim 15, wherein the contacting is effected in vitro.

19. The method of claim 15, wherein the contacting is effected in vivo.

20. The method of claim 15, wherein the regulatory activity is stimulation of cell proliferation or cell growth inhibition.

21. The method of claim 15, wherein the regulatory activity is promotion of extracellular matrix production.

22. The method of claim 15, wherein the regulatory activity is regulation of immune function.

23. The method of claim 15, wherein the polypeptide is a purified human endoglin-derived polypeptide comprising a disulfide-linked homodimer of about 80 to about 95 kDa subunits or a biologically active fragment thereof.

24. The method of claim 23, wherein the purified polypeptide is purified and isolated human endoglin-derived protein substantially free of other human proteins.

25. The method of claim 15, wherein the polypeptide is a soluble endoglin-derived polypeptide.

26. A method of treating a pathologic condition caused by TGF-β regulated cell growth stimulation which comprises administering to a subject an effective amount of the pharmaceutical composition of claim 13 or 14 to bind to the TGF-ß thereby treating the pathologic condition caused by cell growth stimulation.

27. The method of claim 26, wherein TGF-ß is TGF-ß1.

28. The method of claim 26, wherein TGF-ß is TGF-ß3.

29. The method of claim 26, wherein the subject is a human patient.

30. A method of treating a pathologic condition caused by TGF-β regulated inhibition of cell growth which comprises administering to a subject an effective amount of the pharmaceutical composition of claim 13 or 14 to bind to the TGF-ß, thereby treating the pathologic condition caused by inhibition of cell growth.

31. The method of claim 30, wherein TGF-ß is TGF-ß1.

32. The method of claim 30, wherein TGF-ß is TGF-ß3.

33. The method of claim 30, wherein the pathologic condition is ulceration or immune suppression.

34. The method of claim 30, wherein the subject is a human patient.

35. The method of treating a pathologic condition caused by TGF-ß regulated promotion of extracellular matrix accumulation which comprises administering to a subject an effective amount of the pharmaceutical composition of claim 12 or 13 to bind TGF-ß thereby treating the pathologic condition caused by promotion of extracellular matrix.

36. The method of claim 35, wherein the subject is a human patient.

37. The method of claim 35, wherein the pathologic condition is inflammation, rheumatoid arthritis, inflamed skin lesions, scar tissue formation, lung fibrosis, liver fibrosis, atherosclerosis or glomerulonephritis.

38. A method of inhibiting the activity of endoglin which comprises contacting endoglin with an effective amount of a polypeptide capable of binding to endoglin to bind endoglin, thereby inhibiting the activity of endoglin.

39. The method of claim 38, wherein the polypeptide is a TGF-ß.

40. The method of claim 38, wherein the polypeptide has an amino acid sequence corresponding to a fragment of a TGF-ß having the ability to bind endoglin.

41. The method of claim 39, wherein the TGF-ß is TGF-ß1 or TGF-ß3.

42. The method of claim 40, wherein the TGF-ß is TGF-ß1 or TGF-ß3.

43. The method of claim 38, wherein the polypeptide is an anti-endoglin antibody.

44. The method of claim 43, wherein the antibody is a monoclonal antibody.

45. The method of claim 38, wherein the contacting is effected in vivo.

46. The method of claim 38, wherein the contacting is effected in vitro.

FIGURE 1

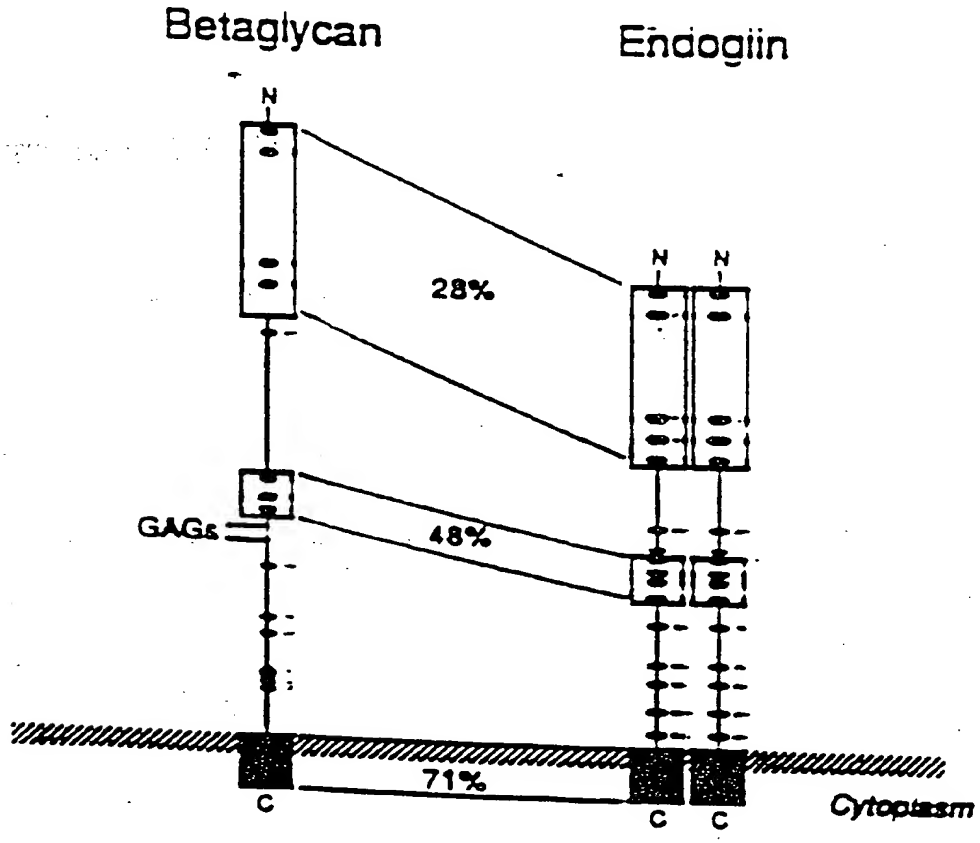


FIGURE 2

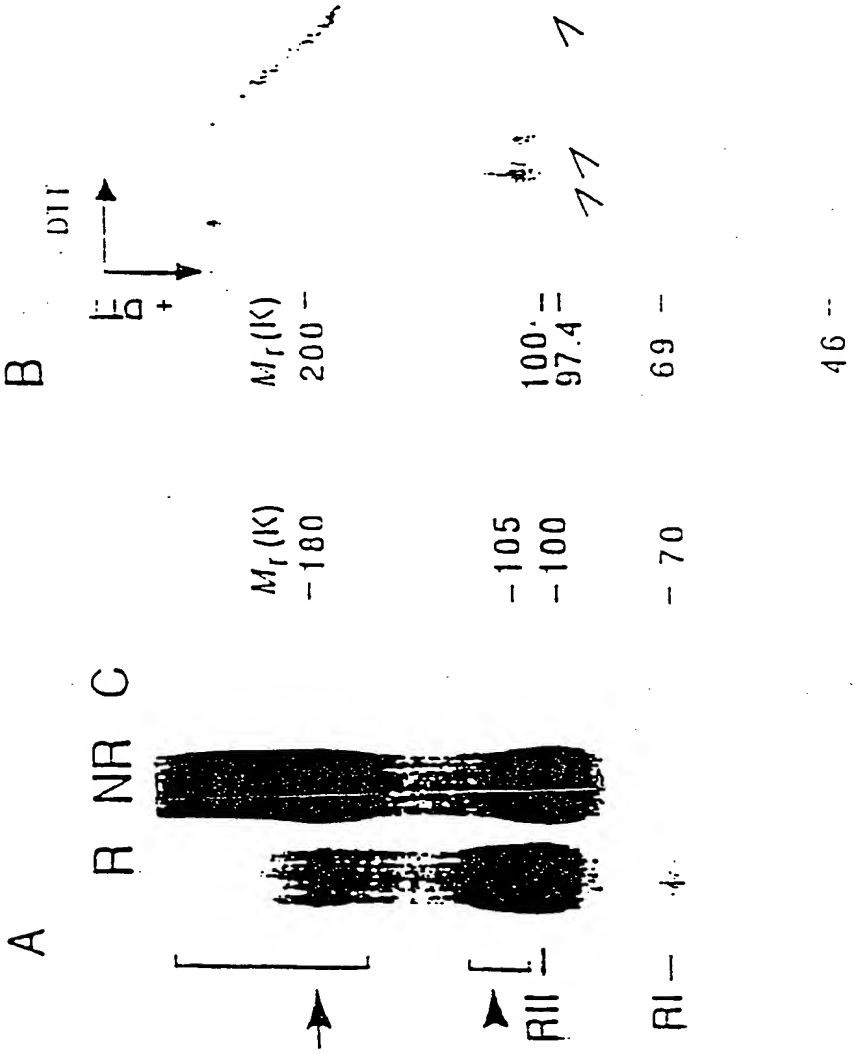


FIGURE 3

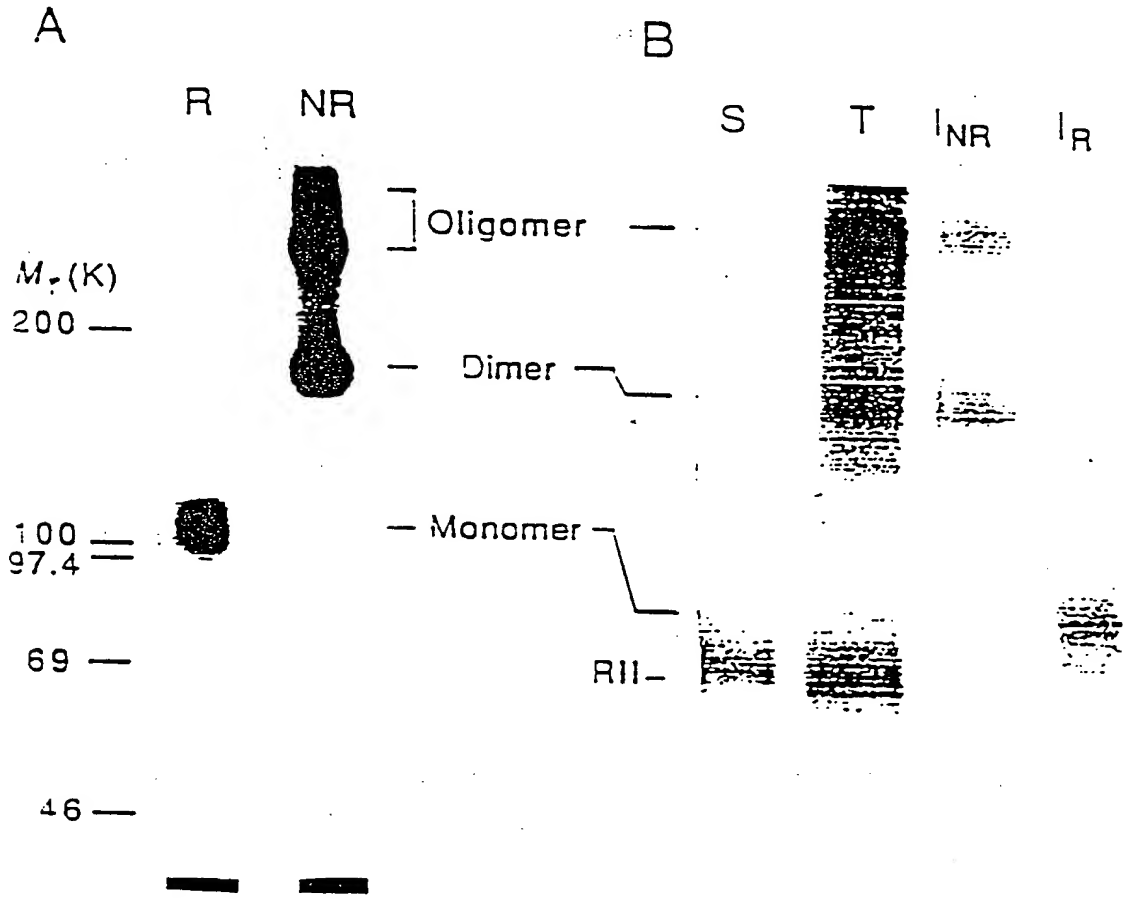

A
R
NR
B
S
T
I_{NR}
I_R
Oligomer
Dimer
Monomer
RII
M_r(K)
200
100
97.4
69
46

FIGURE 4

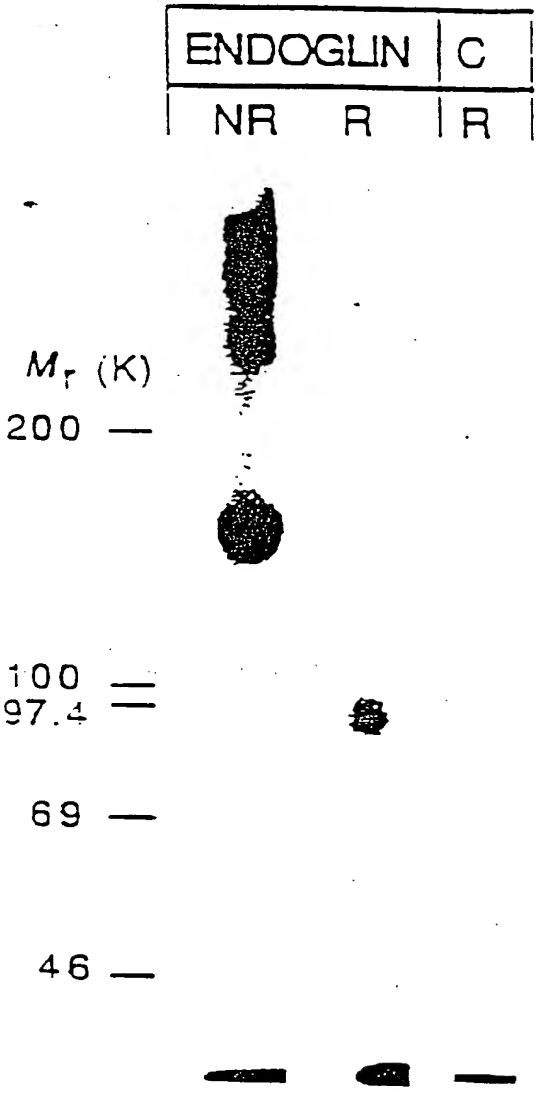
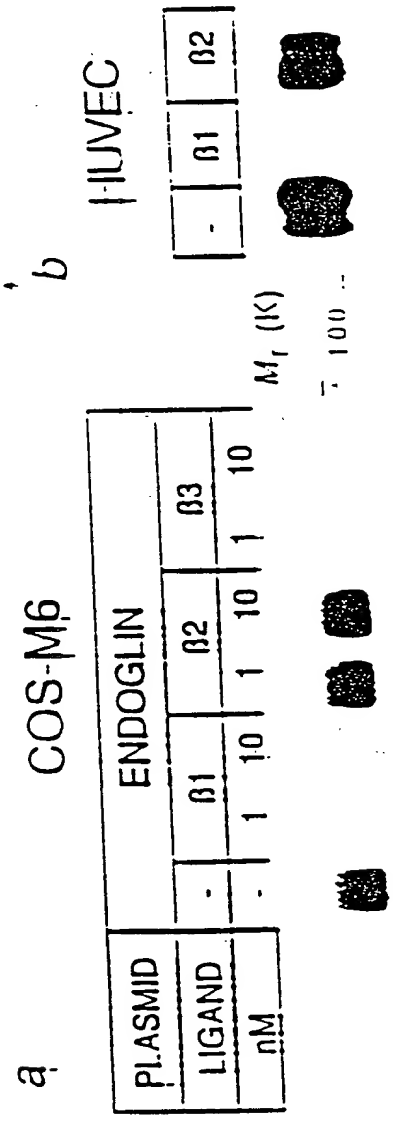
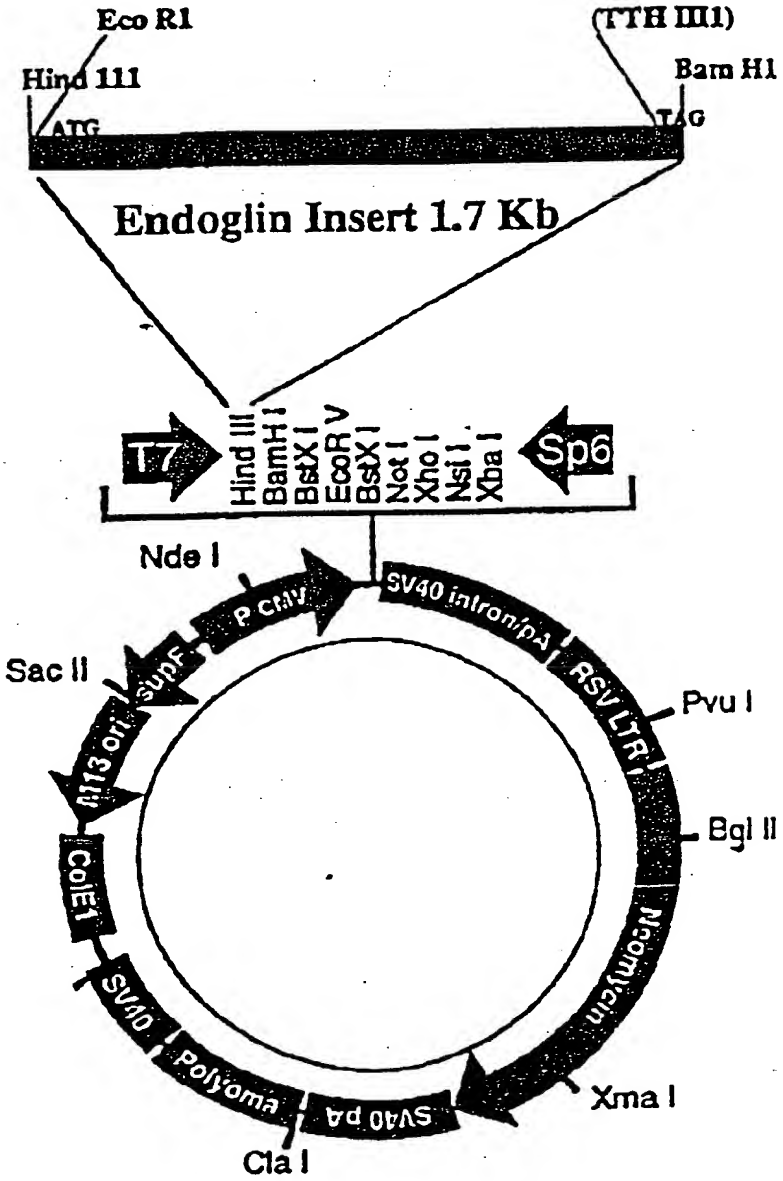


FIGURE 5

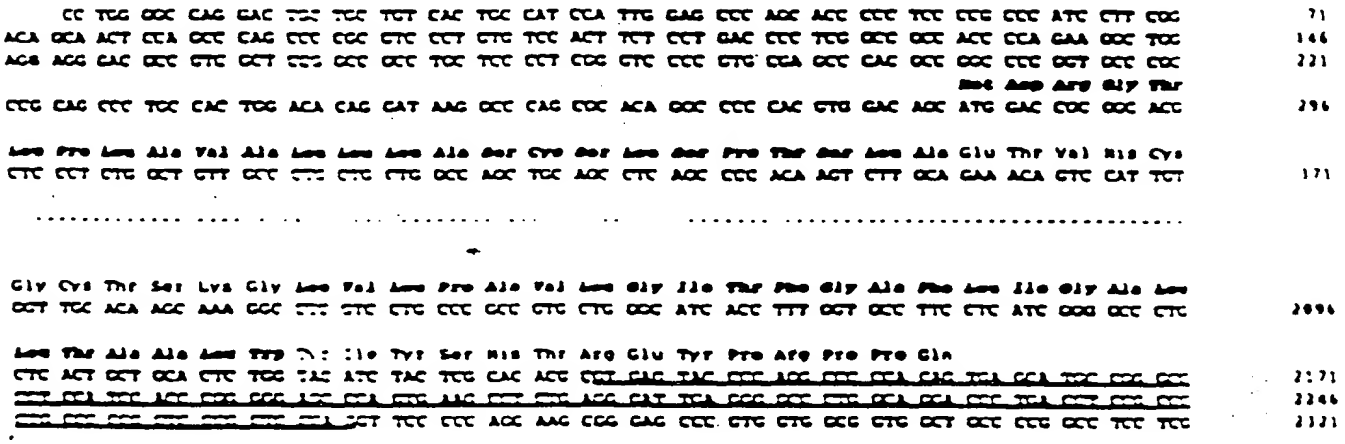

a
COS-M6
PLASMID
LIGAND
nM
ENDOGLIN
-
β1
β2
β3
1
10
b
HUVEC
M_r (K)
100

pcNeoSolEND
8.7 Kb



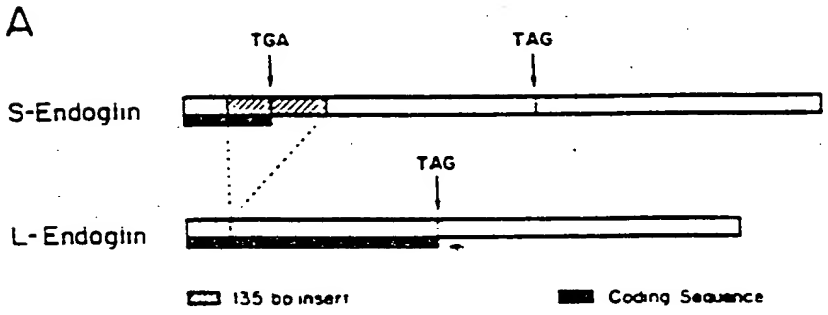
Vector pcDNAI/Neo
7.0 Kb

FIGURE 7



THIS PAGE BLANK (USPTO)

FIGURE 8


A
TGA
TAG
S-Endoglin
TAG
L-Endoglin
135 bp insert
Coding Sequence

B

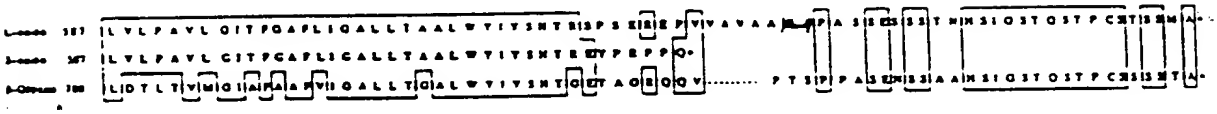
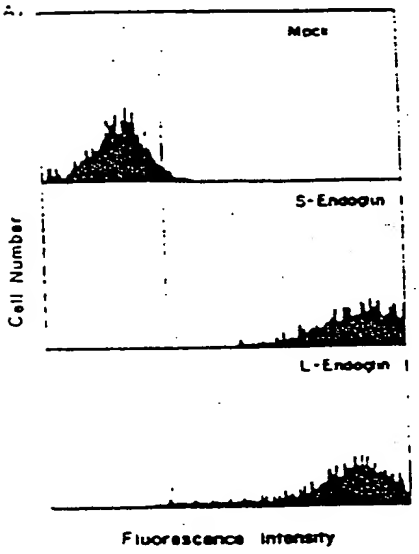
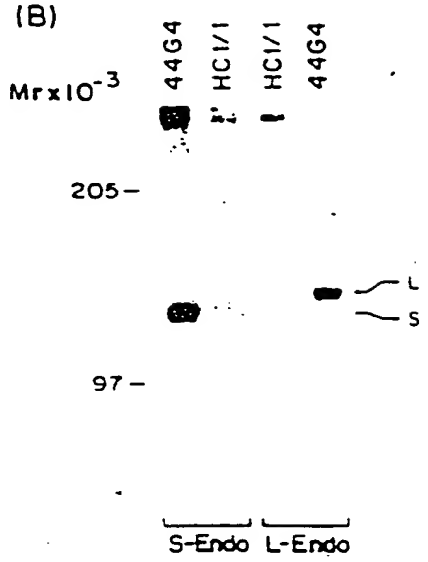
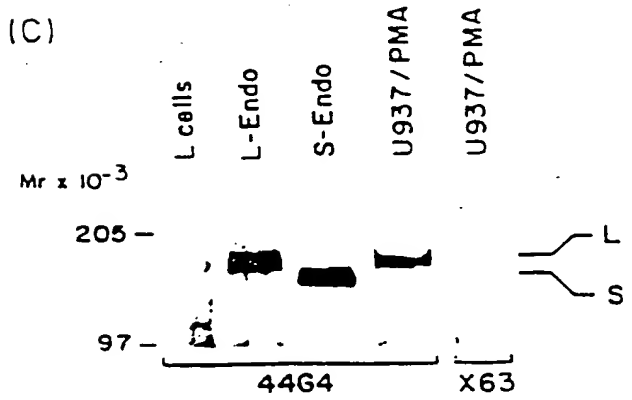
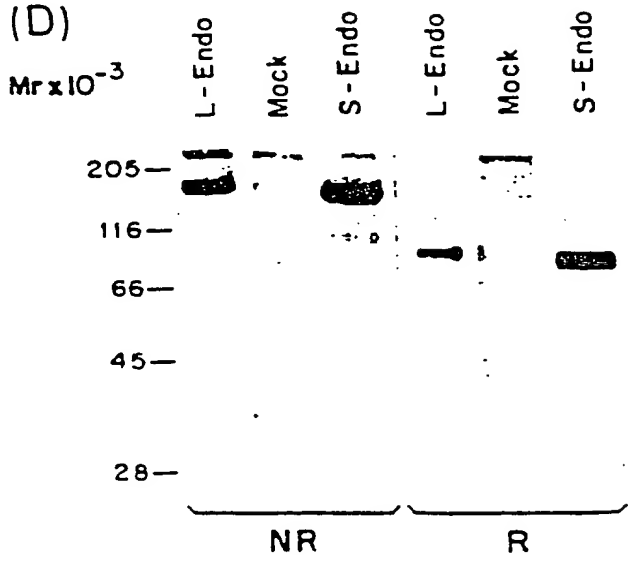


FIGURE 9


Mock
S-Endoglin
L-Endoglin
Cell Number
Fluorescence Intensity


(B)
Mr x10-3
44G4
HC1/1
HC1/1
44G4
205
97
L
S
S-Endo
L-Endo


(C)
L cells
L-Endo
S-Endo
U937/PMA
U937/PMA
Mr x 10-3
205
97
L
S
44G4
X63


(D)
Mr x10-3
L-Endo
Mock
S-Endo
L-Endo
Mock
S-Endo
205
116
66
45
28
NR
R

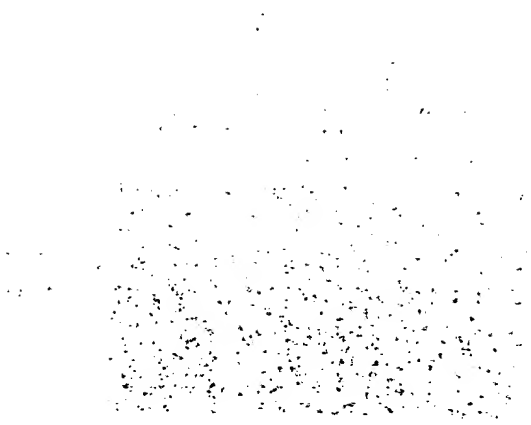


FIGURE 10

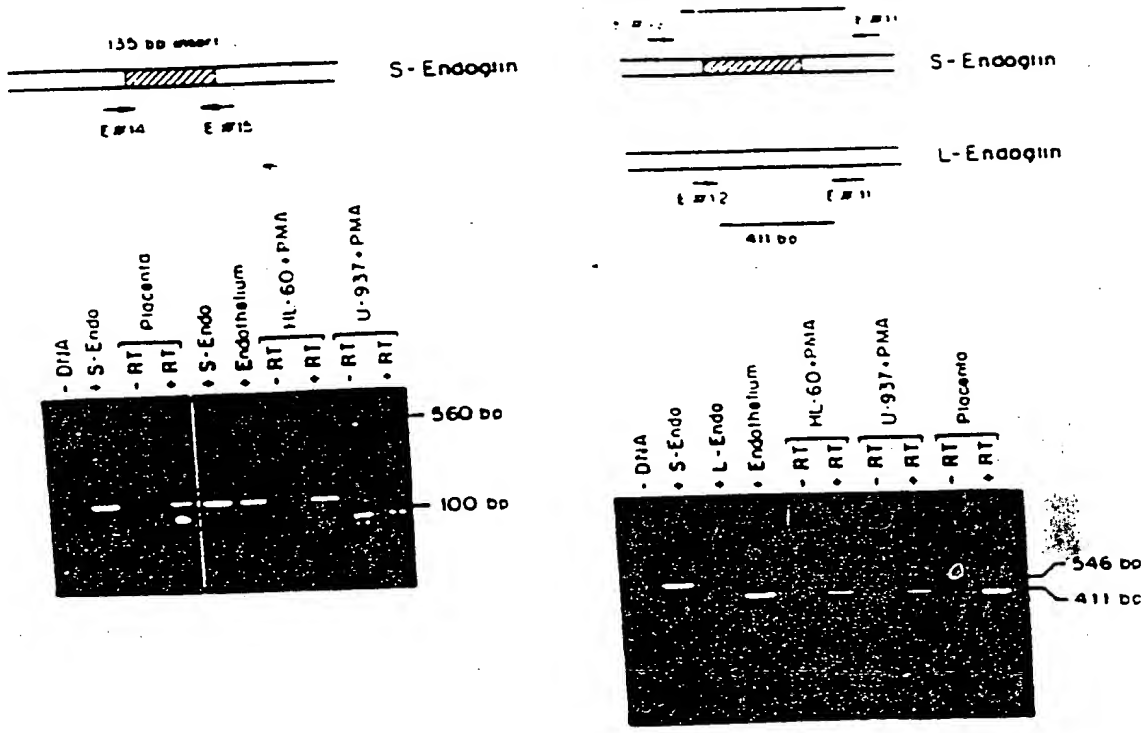
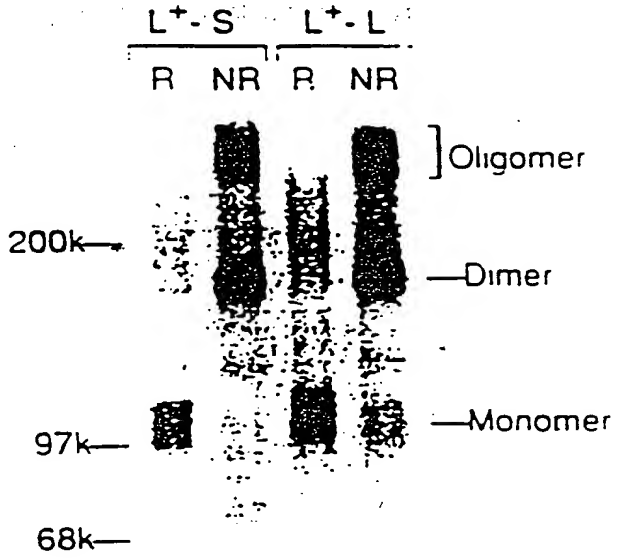

S-Endoglin
L-Endoglin
560 bp
100 bp
546 bp
411 bp
- DNA
+ S-Endo
+ L-Endo
- RT
+ RT
Placenta
+ Endothelium
HL-60 + PMA
U-937 + PMA

FIGURE 11



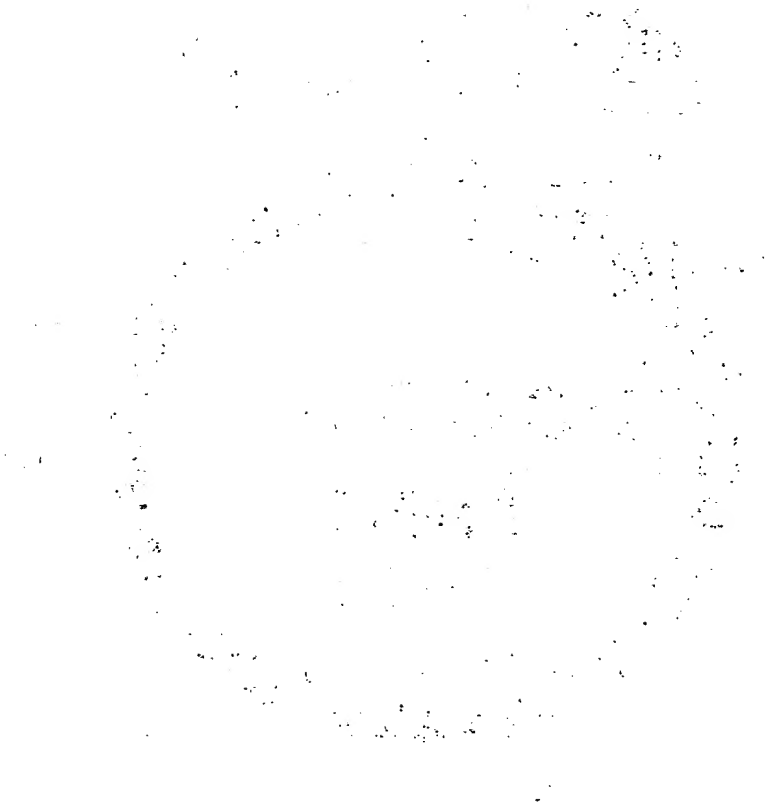


FIGURE 12

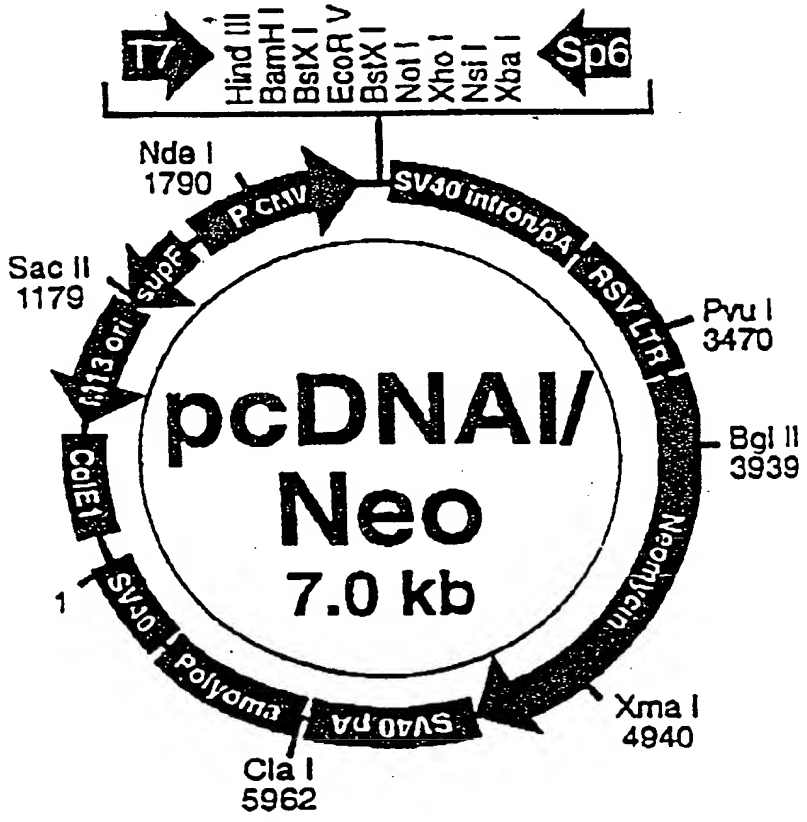
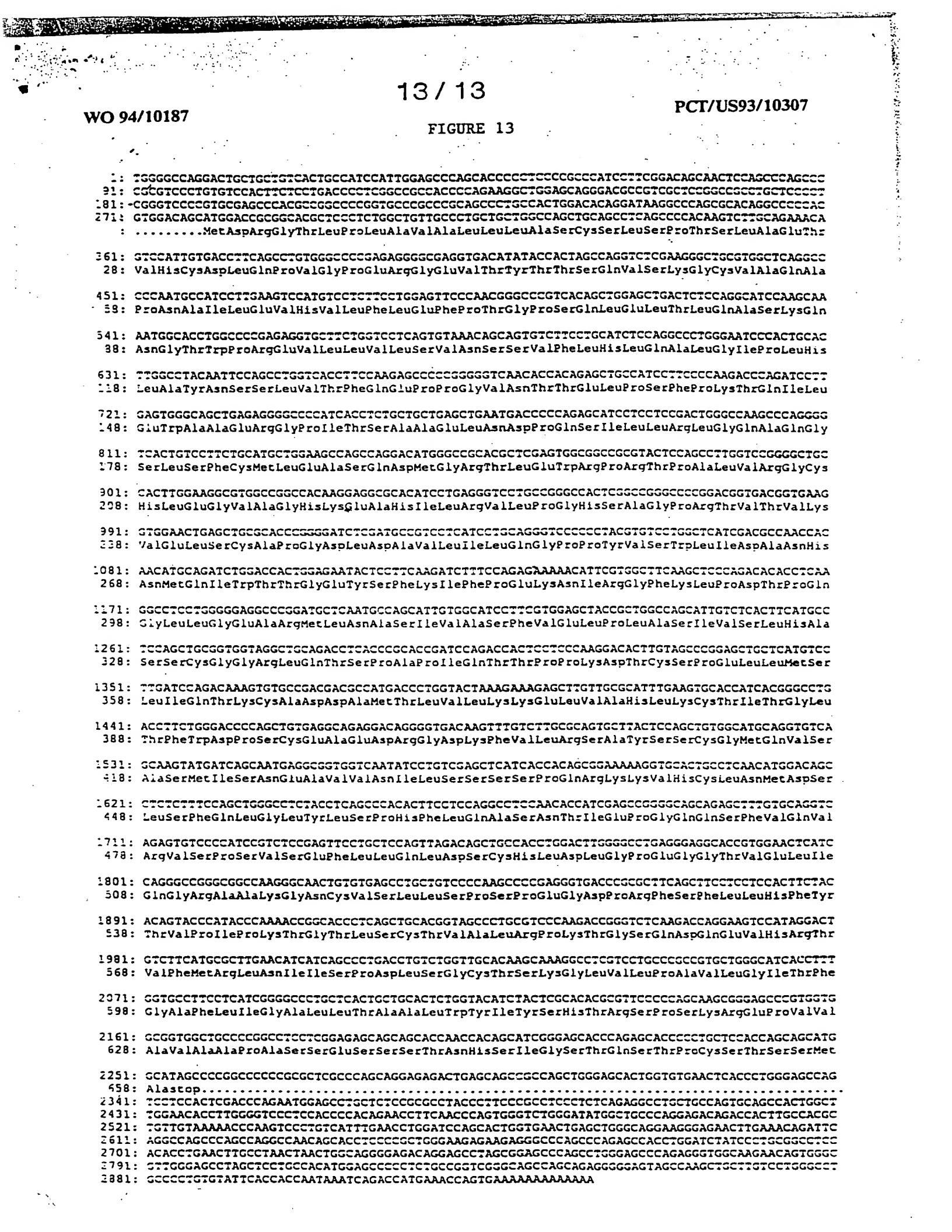

T7
Hind III
BamH I
BstX I
EcoR V
BstX I
Not I
Xho I
Nsi I
Xba I
Sp6
Nde I
1790
P CMV
SV40 intron/pA
RSV LTR
Pvu I
3470
Sac II
1179
SupF
f1 ori
pcDNAI/
Neo
7.0 kb
Bgl II
3939
ColE1
1
SV40
Polyoma
SV40 pA
Neomycin
Xma I
4940
Cla I
5962

FIGURE 13


1: TGGGGCCAGGACTGCTGCTGTCACTGCCATCCATTGGAGCCCAGCACCCCCTCCCCGCCCATCCTTCGGACAGCAACTCCAGCCCAGCCC
91: CGCGTCCCTGTGTCCACTTCTCCTGACCCCTCGGCCGCCACCCCAGAAGGCTGGAGCAGGGACGCCGTCGCTCCGGCCGCCTGCTCCCCT
181: -CGGGTCCCCGTGCGAGCCCACGCCGGCCCCGGTGCCCGCCCGCAGCCCTGCCACTGGACACAGGATAAGGCCCAGCGCACAGGCCCCCAC
271: GTGGACAGCATGGACCGCGGCACGCTCCCTCTGGCTGTTGCCCTGCTGCTGGCCAGCTGCAGCCTCAGCCCCACAAGTCTTGCAGAAACA
: .....MetAspArgGlyThrLeuProLeuAlaValAlaLeuLeuLeuAlaSerCysSerLeuSerProThrSerLeuAlaGluThr
361: GTCCATTGTGACCTTCAGCCTGTGGGCCCCGAGAGGGGCGAGGTGACATATACCACTAGCCAGGTCTCGAAGGGCTGCGTGGCTCAGGCC
28: ValHisCysAspLeuGlnProValGlyProGluArgGlyGluValThrTyrThrThrSerGlnValSerLysGlyCysValAlaGlnAla
451: CCCAATGCCATCCTTGAAGTCCATGTCCTCTTCCTGGAGTTCCCAACGGGCCCGTCACAGCTGGAGCTGACTCTCCAGGCATCCAAGCAA
58: ProAsnAlaIleLeuGluValHisValLeuPheLeuGluPheProThrGlyProSerGlnLeuGluLeuThrLeuGlnAlaSerLysGln
541: AATGGCACCTGGCCCCGAGAGGTGCTTCTGGTCCTCAGTGTAAACAGCAGTGTCTTCCTGCATCTCCAGGCCCTGGGAATCCCACTGCAC
88: AsnGlyThrTrpProArgGluValLeuLeuValLeuSerValAsnSerSerValPheLeuHisLeuGlnAlaLeuGlyIleProLeuHis
631: TTGGCCTACAATTCCAGCCTGGTCACCTTCCAAGAGCCCCCGGGGGTCAACACCACAGAGCTGCCATCCTTCCCCAAGACCCAGATCCTT
118: LeuAlaTyrAsnSerSerLeuValThrPheGlnGluProProGlyValAsnThrThrGluLeuProSerPheProLysThrGlnIleLeu
721: GAGTGGGCAGCTGAGAGGGGCCCCATCACCTCTGCTGCTGAGCTGAATGACCCCCAGAGCATCCTCCTCCGACTGGGCCAAGCCCAGGGG
148: GluTrpAlaAlaGluArgGlyProIleThrSerAlaAlaGluLeuAsnAspProGlnSerIleLeuLeuArgLeuGlyGlnAlaGlnGly
811: TCACTGTCCTTCTGCATGCTGGAAGCCAGCCAGGACATGGGCCGCACGCTCGAGTGGCGGCCGCGTACTCCAGCCTTGGTCCGGGGCTGC
178: SerLeuSerPheCysMetLeuGluAlaSerGlnAspMetGlyArgThrLeuGluTrpArgProArgThrProAlaLeuValArgGlyCys
901: CACTTGGAAGGCGTGGCCGGCCACAAGGAGGCGCACATCCTGAGGGTCCTGCCGGGCCACTCGGCCGGGCCCCGGACGGTGACGGTGAAG
208: HisLeuGluGlyValAlaGlyHisLysGluAlaHisIleLeuArgValLeuProGlyHisSerAlaGlyProArgThrValThrValLys
991: GTGGAACTGAGCTGCGCACCCGGGGATCTCGATGCCGTCCTCATCCTGCAGGGTCCCCCCTACGTGTCCTGGCTCATCGACGCCAACCAC
238: ValGluLeuSerCysAlaProGlyAspLeuAspAlaValLeuIleLeuGlnGlyProProTyrValSerTrpLeuIleAspAlaAsnHis
1081: AACATGCAGATCTGGACCACTGGAGAATACTCCTTCAAGATCTTTCCAGAGAAAAACATTCGTGGCTTCAAGCTCCCAGACACACCTCAA
268: AsnMetGlnIleTrpThrThrGlyGluTyrSerPheLysIlePheProGluLysAsnIleArgGlyPheLysLeuProAspThrProGln
1171: GGCCTCCTGGGGGAGGCCCGGATGCTCAATGCCAGCATTGTGGCATCCTTCGTGGAGCTACCGCTGGCCAGCATTGTCTCACTTCATGCC
298: GlyLeuLeuGlyGluAlaArgMetLeuAsnAlaSerIleValAlaSerPheValGluLeuProLeuAlaSerIleValSerLeuHisAla
1261: TCCAGCTGCGGTGGTAGGCTGCAGACCTCACCCGCACCGATCCAGACCACTCCTCCCAAGGACACTTGTAGCCCGGAGCTGCTCATGTCC
328: SerSerCysGlyGlyArgLeuGlnThrSerProAlaProIleGlnThrThrProProLysAspThrCysSerProGluLeuLeuMetSer
1351: TTGATCCAGACAAAGTGTGCCGACGACGCCATGACCCTGGTACTAAAGAAAGAGCTTGTTGCGCATTTGAAGTGCACCATCACGGGCCTG
358: LeuIleGlnThrLysCysAlaAspAspAlaMetThrLeuValLeuLysLysGluLeuValAlaHisLeuLysCysThrIleThrGlyLeu
1441: ACCTTCTGGGACCCCAGCTGTGAGGCAGAGGACAGGGGTGACAAGTTTGTCTTGCGCAGTGCTTACTCCAGCTGTGGCATGCAGGTGTCA
388: ThrPheTrpAspProSerCysGluAlaGluAspArgGlyAspLysPheValLeuArgSerAlaTyrSerSerCysGlyMetGlnValSer
1531: GCAAGTATGATCAGCAATGAGGCGGTGGTCAATATCCTGTCGAGCTCATCACCACAGCGGAAAAAGGTGCACTGCCTCAACATGGACAGC
418: AlaSerMetIleSerAsnGluAlaValValAsnIleLeuSerSerSerSerProGlnArgLysLysValHisCysLeuAsnMetAspSer
1621: CTCTCTTTCCAGCTGGGCCTCTACCTCAGCCCACACTTCCTCCAGGCCTCCAACACCATCGAGCCGGGGCAGCAGAGCTTTGTGCAGGTC
448: LeuSerPheGlnLeuGlyLeuTyrLeuSerProHisPheLeuGlnAlaSerAsnThrIleGluProGlyGlnGlnSerPheValGlnVal
1711: AGAGTGTCCCCATCCGTCTCCGAGTTCCTGCTCCAGTTAGACAGCTGCCACCTGGACTTGGGGCCTGAGGGAGGCACCGTGGAACTCATC
478: ArgValSerProSerValSerGluPheLeuLeuGlnLeuAspSerCysHisLeuAspLeuGlyProGluGlyGlyThrValGluLeuIle
1801: CAGGGCCGGGCGGCCAAGGGCAACTGTGTGAGCCTGCTGTCCCCAAGCCCCGAGGGTGACCCGCGCTTCAGCTTCCTCCTCCACTTCTAC
508: GlnGlyArgAlaAlaLysGlyAsnCysValSerLeuLeuSerProSerProGluGlyAspProArgPheSerPheLeuLeuHisPheTyr
1891: ACAGTACCCATACCCAAAACCGGCACCCTCAGCTGCACGGTAGCCCTGCGTCCCAAGACCGGGTCTCAAGACCAGGAAGTCCATAGGACT
538: ThrValProIleProLysThrGlyThrLeuSerCysThrValAlaLeuArgProLysThrGlySerGlnAspGlnGluValHisArgThr
1981: GTCTTCATGCGCTTGAACATCATCAGCCCTGACCTGTCTGGTTGCACAAGCAAAGGCCTCGTCCTGCCCCGCCGTGCTGGGCATCACCTTT
568: ValPheMetArgLeuAsnIleIleSerProAspLeuSerGlyCysThrSerLysGlyLeuValLeuProAlaValLeuGlyIleThrPhe
2071: GGTGCCTTCCTCATCGGGGCCCTGCTCACTGCTGCACTCTGGTACATCTACTCGCACACGCGTTCCCCCAGCAAGCGGGAGCCCGTGGTG
598: GlyAlaPheLeuIleGlyAlaLeuLeuThrAlaAlaLeuTrpTyrIleTyrSerHisThrArgSerProSerLysArgGluProValVal
2161: GCGGTGGCTGCCCCGGCCTCCTCGGAGAGCAGCAGCACCAACCACAGCATCGGGAGCACCCAGAGCACCCCCTGCTCCACCAGCAGCATG
628: AlaValAlaAlaProAlaSerSerGluSerSerSerThrAsnHisSerIleGlySerThrGlnSerThrProCysSerThrSerSerMet
2251: GCATAGCCCCGGCCCCCCCGCGCTCGCCCAGCAGGAGAGACTGAGCAGCCGCCAGCTGGGAGCACTGGTGTGAACTCACCCTGGGAGCCAG
658: Alastop.....
2341: TCCTCCACTCGACCCAGAATGGAGCCTGCTCTCCGCGCCTACCCTTCCCGCCTCCCTCTCAGAGGCCTGCTGCCAGTGCAGCCACTGGCT
2431: TGGAACACCTTGGGGTCCCTCCACCCCACAGAACCTTCAACCCAGTGGGTCTGGGATATGGCTGCCCAGGAGACAGACCACTTGCCACGC
2521: TGTTGTAAAAACCCAAGTCCCTGTCATTTGAACCTGGATCCAGCACTGGTGAACTGAGCTGGGCAGGAAGGGAGAACTTGAAACAGATTC
2611: AGGCCAGCCCAGCCAGGCCAACAGCACCTCCCCGCTGGGAAGAGAAGAGGGCCCAGCCCAGAGCCACCTGGATCTATCCCTGCGGCCTCC
2701: ACACCTGAACTTGCCTAACTAACTGGCAGGGGAGACAGGAGCCTAGCGGAGCCCAGCCTGGGAGCCCAGAGGGTGGCAAGAACAGTGGGC
2791: GTTGGGAGCCTAGCTCCTGCCACATGGAGCCCCCTCTGCCGGTCGGGCAGCCAGCAGAGGGGGAGTAGCCAAGCTGCTTGTCCTGGGCCT
2881: GCCCCTGTGTATTCACCACCAATAAATCAGACCATGAAACCAGTGAAAAAAAAAAAAA